



(19)

Europäisches Patentamt
European Patent Office
Office européen des brevets

(11)

EP 0 982 401 A3

(12)

EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
19.04.2000 Bulletin 2000/16

(51) Int. Cl. 7:
C12N 15/31, C07K 14/40,
A61K 31/70, A61K 38/16,
C07K 16/14, G01N 33/50,
C12Q 1/68

(43) Date of publication A2:
01.03.2000 Bulletin 2000/09

(21) Application number: 98310694.9

(22) Date of filing: 23.12.1998

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 14.08.1998 GB 9817796

(71) Applicant: JANSSEN PHARMACEUTICA N.V.
2340 Beerse (BE)

(72) Inventors:
, Contreras, Roland Henri
9000 Gent (BE)
, Nelissen, Bart

2340 Beerse (BE)
, De Backer, Marianne Denise
2340 Beerse (BE)
, Luyten, Walter Herman Maria Louis
2340 Beerse (BE)
, Vlaene, Jasmine Elsa
9000 Gent (BE)
, Logghe, Marc George
9000 Gent (BE)

(74) Representative:
Baldock, Sharon Claire et al
BOULT WADE TENNANT,
27 Furnival Street
London EC4A 1PQ (GB)

(54) Drug targets in *Candida albicans*

(57) Nucleic acid molecules encoding polypeptides that are critical for survival and growth of the yeast *Candida albicans* are disclosed. Also provided are methods of identifying compounds which selectively modulate expression or activity of such polypeptides comprising the steps of (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid

molecule according to the invention which mutation results in overexpression or underexpression of said polypeptides in addition to contacting one or more wild type *Candida albicans* cells with said compound, and (b) monitoring the growth and/or activity of said mutated cell compared to said wild type; wherein differential growth or activity of said one or more mutated *Candida* cells is indicative of selective action of said compound on a polypeptide or another polypeptide in the same or a parallel pathway.

EP 0 982 401 A3

ATTORNEY DOCKET NUMBER: 10182-016-999
SERIAL NUMBER: 10/032,585
REFERENCE: CJ



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

which under Rule 45 of the European Patent Convention EP 98 31 0694
shall be considered, for the purposes of subsequent
proceedings, as the European search report

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	DALY S ET AL: "Isolation and characterization of a gene encoding alpha-tubulin from <i>Candida albicans</i> " GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, GB, ELSEVIER SCIENCE PUBLISHERS, BARKING, vol. 187, no. 2, page 151-158 XP004093273 ISSN: 0378-1119 * the whole document * ---		C12N15/31 C07K14/40 A61K31/70 A61K38/16 C07K16/14 G01N33/50 C12Q1/68
A	WO 97 36925 A (SCRIPTGEN PHARM INC ; HARVARD COLLEGE (US)) 9 October 1997 (1997-10-09) * the whole document *		
A	WO 97 37230 A (BRADLEY JOHN; WOBBE C RICHARD; BURATOWSKI STEPHEN) 9 October 1997 (1997-10-09) * the whole document *		
A	WO 96 36707 A (UNIV ROMA ; IST SUPERIORE SANITA (IT); CASSONE ANTONIO (IT); VALLE) 21 November 1996 (1996-11-21) * the whole document *		
INCOMPLETE SEARCH			TECHNICAL FIELDS SEARCHED (Int.Cl.7)
			C12N C07K A61K G01N C12Q
<p>The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims.</p> <p>Claims searched completely :</p> <p>Claims searched incompletely :</p> <p>Claims not searched :</p> <p>Reason for the limitation of the search: see sheet C</p>			
Place of search		Date of completion of the search	Examiner
THE HAGUE		19 November 1999	Smalt, R
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons S : member of the same patent family, corresponding document</p>			



CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing more than ten claims.

Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid, namely claim(s):

No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

see sheet B

All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.

As all searchable claims could be searched without effort justifying an additional fee, the Search Division did not invite payment of any additional fee.

Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:

None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:

1,4-7,9,11-20,30,31 partially



The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. Claims: Invention 1: claims 1,4-7,9,11-20,30,31 partially

Nucleic acid molecule comprising seq.ID.1 or capable of hybridizing thereto, expression vector comprising said nucleic acid, use of said vector for preparation of medicament or pharmaceutical composition, *C. albicans* cell comprising an induced mutation in said DNA sequence, oligonucleotides comprising 10-50 nt of said nucleic acid sequence, and method for identifying compounds which modulate expression of said nucleic acid.

2. Claims: Inventions 2-41: claims 1,4-7,9,11-20,30, 31 partially, and 2,3,8,10,32, 33 partially as applicable

As invention 1, but limited to the respective nucleic acid sequences 2,3,5,6,8,9,10,11,13,15,16,18,20,21,23,25,26,27, 28,29,31,35,37,39,41,43,45,47,49,51,53,55,57,59,61,63,65,67, 69, and 71, and polypeptide sequences corresponding to said nucleic acid sequences in as far as they are provided, whereby invention 2 is limited to seq.ID.2, invention 3 is limited to seq.ID.3 and its translated polypeptide seq.ID.4,, and invention 41 is limited to seq.ID.71 and its translated polypeptide sequence seq.ID.72.

In as far as a polypeptide sequence, translated from the ORF of a corresponding nucleic acid sequence is provided, the polypeptide encoded by the corresponding nucleic acid sequence and their use in the preparation of a medicament, and antibodies against said polypeptide is also considered part of the respective invention.

3. Claims: Invention 42: claim 25-29

Method for identifying DNA sequences from a cell or organism, which encode polypeptides which are critical for growth and survival for said cell or organism, comprising screening a library of nucleic acids using a vector that either integrates into the genome of said cell or organism, or that permits expression of antisense RNA, and selecting growth-impaired cells or organisms. Plasmids pGAL1PSiST-1 and pGAL1PNiST-1, used in said method.



Claim(s) not searched:
21-24

Reason for the limitation of the search:

Claims 21-24 refer to a compound identifiable with a method, without giving a true technical characterization of the compound. Moreover, no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 83 and 84 EPC). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

EP 0 982 401 A3
ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.

EP 98 31 0694

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

19-11-1999

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
WO 9736925	A	09-10-1997	CA EP	2250129 A 0904289 A	09-10-1997 31-03-1999
WO 9737230	A	09-10-1997	US CA EP	5863762 A 2250121 A 0894269 A	26-01-1999 09-10-1997 03-02-1999
WO 9636707	A	21-11-1996	IT AU EP	RM950314 A 5777696 A 0826040 A	18-11-1996 29-11-1996 04-03-1998

Description

[0001] The present invention is concerned with the identification of genes or functional fragments thereof from *Candida albicans* which are critical for growth and cell division and which genes may be used as selective drug targets to treat *Candida albicans* associated infections. Novel nucleic acid sequences from *Candida albicans* are also provided and which encode the polypeptides which are critical for growth of *Candida albicans*.

[0002] Opportunistic infections in immunocompromised hosts represent an increasingly common cause of mortality and morbidity. *Candida* species are among the most commonly identified fungal pathogens associated with such opportunistic infections, with *Candida albicans* being the most common species. Such fungal infections are thus problematical in, for example, AIDS populations in addition to normal healthy women where *Candida albicans* yeasts represent the most common cause of vulvovaginitis.

[0003] Although compounds do exist for treating such disorders, such as for example, amphotericin, these drugs are generally limited in their treatment because of their toxicity and side effects. Therefore, there exists a need for new compounds which may be used to treat *Candida* associated infections in addition to compounds which are selective in their action against *Candida albicans*.

[0004] Classical approaches for identifying anti-fungal compounds have relied almost exclusively on inhibition of fungal or yeast growth as an endpoint. Libraries of natural products, semi-synthetic, or synthetic chemicals are screened for their ability to kill or arrest growth of the target pathogen or a related nonpathogenic model organism. These tests are cumbersome and provide no information about a compounds mechanism of action. The promising lead compounds that emerge from such screens must then be tested for possible host-toxicity and detailed mechanism of action studies must subsequently be conducted to identify the affected molecular target.

[0005] The present inventors have now identified a range of nucleic acid sequences from *Candida albicans* which encode polypeptides which are critical for its survival and growth. These sequences represent novel targets which can be incorporated into an assay to selectively identify compounds capable of inhibiting expression of such polypeptides and their potential use in alleviating diseases or conditions associated with *Candida albicans* infection.

[0006] Therefore, according to a first aspect of the invention there is provided a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides in Sequence ID Numbers 1 to 3, 5, 6, 8 to 11, 13, 15, 16, 18, 20, 21, 23, 25 to 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69 and 71.

[0007] A further aspect of the invention comprises a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of Sequence ID Numbers 1, 28, 35, 37 and 39 and fragments or derivatives of said nucleic acid molecules.

[0008] Also provided by the present invention is a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which polypeptide has an amino acid sequence according to the sequence of any of Sequence ID Numbers 4, 7, 12, 14, 17, 19, 22, 24, 30, 32 to 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 and 72.

[0009] Letters utilised in the nucleic acid sequences according to the invention which are not recognisable as letters of the genetic code signify a position in the nucleic acid sequence where one or more of bases A, G, C or T can occupy the nucleotide position. Representative letters used to identify the range of bases which can be used are as follows:

M:	A or C
R:	A or G
W:	A or T
S:	C or G
Y:	C or T
K:	G or T
V:	A or C or G
H:	A or C or T
D:	A or G or T
B:	C or G or T
N:	G or A or T or C

[0010] In one embodiment of the above identified aspects of the invention the nucleic acid may comprise a mRNA molecule or alternatively a DNA and preferably a cDNA molecule.

[0011] Also provided by the present invention is a nucleic acid molecule capable of hybridising to the nucleic acid molecules according to the invention under high stringency conditions.

[0012] Stringency of hybridisation as used herein refers to conditions under which polynucleic acids are stable. The stability of hybrids is reflected in the melting temperature (Tm) of the hybrids. Tm can be approximated by the formula:

$$81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41 (\%G\&C) - 6001/l$$

wherein l is the length of the hybrids in nucleotides. Tm decreases approximately by 1-1.5°C with every 1% decrease in sequence homology.

[0013] The nucleic acid capable of hybridising to nucleic acid molecules according to the invention will generally be at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the

nucleotide sequences according to the invention.

[0014] The DNA molecules according to the invention may, advantageously, be included in a suitable expression vector to express polypeptides encoded therefrom in a suitable host.

[0015] The present invention also comprises within its scope proteins or polypeptides encoded by the nucleic acid molecules according to the invention or a functional equivalent, derivative or bioprecursor thereof.

[0016] Therefore, according to a further aspect of the invention there is provided a polypeptide having an amino acid sequence of any of Sequence ID Numbers 4, 7, 12, 14, 17, 19, 22, 24, 30, 32 to 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 and 72. A polypeptide encoded by the nucleic acid molecule according to the invention is also provided, which polypeptide preferably comprises an amino acid sequence of having the sequence of any of Sequence ID Numbers 4, 7, 12, 14, 17, 19, 22, 24, 30, 32 to 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 and 72.

[0017] An expression vector according to the invention includes a vector having a nucleic acid according to the invention operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide according to the invention. Thus, in a further aspect, the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell, transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

[0018] The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of said nucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, ampicillin resistance.

[0019] Polynucleotides according to the invention may be inserted into the vectors described in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense nucleic acids may be produced by synthetic means.

[0020] In accordance with the present invention, a defined nucleic acid includes not only the identical nucleic acid but also any minor base variations including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid residue) due to the degenerate code in conservative amino acid substitutions. The term "nucleic acid sequence" also includes the complementary sequence to any single stranded sequence given regarding base variations.

[0021] The present invention also advantageously provides nucleic acid sequences of at least approximately 10 contiguous nucleotides of a nucleic acid according to the invention and preferably from 10 to 50 nucleotides. These sequences may, advantageously be used as probes or primers to initiate replication, or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with the sample under hybridising conditions and detecting for the presence of any duplex or triplex formation between the probe and any nucleic acid in the sample.

[0022] According to the present invention these probes may be anchored to a solid support. Preferably, they are present on an array so that multiple probes can simultaneously hybridize to a single biological sample. The probes can be spotted onto the array or synthesised *in situ* on the array. (See Lockhart *et al.*, Nature Biotechnology, vol. 14, December 1996 "Expression monitoring by hybridisation to high density oligonucleotide arrays". A single array can contain more than 100, 500 or even 1,000 different probes in discrete locations.

[0023] Advantageously, the nucleic acid sequences, according to the invention may be produced using such recombinant or synthetic means, such as for example using PCR cloning mechanisms which generally involve making a pair of primers, which may be from approximately 10 to 50 nucleotides to a region of the gene which is desired to be cloned, bringing the primers into contact with mRNA, cDNA, or genomic DNA from a human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified region or fragment and recovering the amplified DNA. Generally, such techniques as defined herein are well known in the art, such as described in Sambrook *et al* (Molecular Cloning: a Laboratory Manual, 1989).

[0024] The nucleic acids or oligonucleotides according to the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{39}S , enzyme labels or other protein labels such as biotin or fluorescent markers. Such labels may be added to the nucleic acids or oligonucleotides of the invention and may be detected using known techniques *per se*.

[0025] The polypeptide or protein according to the invention includes all possible amino acid variants encoded by the nucleic acid molecule according to the invention including a polypeptide encoded by said molecule and having conservative amino acid changes. Polypeptides according to the invention further include variants of such sequences, including naturally occurring allelic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 70%, preferably 80 or 90% amino acid homology with the polypeptides encoded by the nucleic acid molecules according to the invention.

[0026] A nucleic acid which is particularly advantageous is one comprising the sequences of nucleotides illustrated in Figures 1 which is specific to *Candida albicans* with no functionally related sequences in other prokaryotic or eukaryotic organism as yet identified from the respective genomic databases.

EP 0 982 401 A2

[0027] Nucleotide sequences according to the invention are particularly advantageous for selective therapeutic targets for treating *Candida albicans* associated infections. For example, an antisense nucleic acid capable of binding to the nucleic acid sequences according to the invention may be used to selectively inhibit expression of the corresponding polypeptides, leading to impaired growth of the *Candida albicans* with reductions of associated illnesses or diseases.

5 [0028] The nucleic acid molecule or the polypeptide according to the invention may be used as a medicament, or in the preparation of a medicament, for treating diseases or conditions associated with *Candida albicans* infection.

[0029] Advantageously, the nucleic acid molecule or the polypeptide according to the invention may be provided in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

10 [0030] Antibodies to the protein or polypeptide of the present invention may, advantageously, be prepared by techniques which are known in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with the polypeptide according to the invention or an epitope thereof and recovering immune serum. Monoclonal antibodies may be prepared according to known techniques such as described by Kohler R. and Milstein C., *Nature* (1975) 256, 495-497.

15 [0031] Antibodies according to the invention may also be used in a method of detecting for the presence of a polypeptide according to the invention, which method comprises reacting the antibody with a sample and identifying any protein bound to said antibody. A kit may also be provided for performing said method which comprises an antibody according to the invention and means for reacting the antibody with said sample.

[0032] Proteins which interact with the polypeptide of the invention may be identified by investigating protein-protein interactions using the two-hybrid vector system first proposed by Chien *et al* (1991).

20 [0033] This technique is based on functional reconstitution in vivo of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as a library or the like, encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

25 [0034] An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example the nucleic acids according to the invention. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein. Any interaction between polypeptides encoded by the nucleic acid according to the invention and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter molecule such as β -galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes.

30 [0035] Further provided by the present invention is one or more *Candida albicans* cells comprising an induced mutation in the DNA sequence encoding the polypeptide according to the invention.

35 [0036] A further aspect of the invention provides a method of identifying compounds which selectively inhibit or interfere with the expression, or the functionality of polypeptides expressed from the nucleotides sequences according to the invention or the metabolic pathways in which these polypeptides are involved and which are critical for growth and survival of *Candida albicans*, which method comprises (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to the invention which mutation results in overexpression or underexpression of said polypeptides in addition to one or more wild type *Candida* cells, (b) monitoring the growth and/or activity of said mutated cell compared to said wild type wherein differential growth or activity of said one or more mutated *Candida* cells provides an indication of selective action of said compound on said polypeptide or another polypeptide in the same or a parallel pathway.

40 [0037] Compounds identifiable or identified using the method according to the invention, may advantageously be used as a medicament, or in the preparation of a medicament to treat diseases or conditions associated with *Candida albicans* infection. These compounds may also advantageously be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

45 [0038] A further aspect of the invention provides a method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival, which method comprises (a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide sequences in said cDNA or genomic library, (b) selecting transformants exhibiting impaired growth and determining the nucleotide sequence of the cDNA or genomic sequence from the library included in the vector from said transformant. Preferably, the cell or organism may be any yeast or filamentous fungi, such as, for example, *Saccharomyces cerevisiae*, *Saccharomyces pombe* or *Candida albicans*.

50 [0039] A further aspect of the invention provides a pharmaceutical composition comprising a compound according

EP 0 982 401 A2

to the invention together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

[0040] A further aspect of the invention comprises nucleic acid molecules encoding proteins which are critical for survival and growth of *Candida albicans*, which nucleic acid molecules comprise any of the sequences illustrated in Figures 5 to 29. Polypeptides which are critical for survival and growth of *Candida albicans* are also encompassed within the present invention, and which polypeptides comprise any of the amino acid sequences illustrated in Figures 29 to 39.

[0041] The present invention may be more clearly understood with reference to the accompanying example, which is purely exemplary, with reference to the accompanying drawings wherein:

Figure 1: is a diagrammatic representation of plasmid pGAL1PNiST-1.

Figure 2: is a nucleotide sequence of plasmid pGAL1PNiST-1 of Figure 1.

Figure 3: is a diagrammatic representation of plasmid pGAL1PSiST-1.

Figure 4: is a nucleotide sequence of plasmid pGAL1PSiST-1 of Figure 3.

Figures 5 to 28: illustrate the nucleotide sequences of oligonucleotides encoding polypeptides of previously unknown function isolated from *Candida albicans* which are critical for its survival and growth, according to the invention.

Figures 29 to 39: illustrate the amino acid sequences of polypeptides from *Candida albicans* which are critical for its survival and growth, according to the invention.

Example 1

Identification of novel drug targets in *C. albicans* by anti-sense and disruptive integration

[0042] The principle of the approach is based on the fact that when a particular *C. albicans* mRNA is inhibited by producing the complementary anti-sense RNA, the corresponding protein will decrease. If this protein is critical for growth or survival, the cell producing the anti-sense RNA will grow more slowly or will die.

[0043] Since anti-sense inhibition occurs at mRNA level, the gene copy number is irrelevant, thus allowing applications of the strategy even in diploid organisms.

[0044] Anti-sense RNA is endogenously produced from an integrative or episomal plasmid with an inducible promoter; induction of the promoter leads to the production of a RNA encoded by the insert of the plasmid. This insert will differ from one plasmid to another in the library. The inserts will be derived from genomic DNA fragments or from cDNA to cover-to-the extent possible- the entire genome.

[0045] The vector is a proprietary vector allowing integration by homologous recombination at either the homologous insert or promoter sequence in the *Candida* genome. After introducing plasmids from cDNA or genomic libraries into *C. albicans*, transformants are screened for impaired growth after promoter (& thus anti-sense) induction in the presence of lithium acetate. Lithium acetate prolongs the G1 phase and thus allows anti-sense to act during a prolonged period of time during the cell cycle. Transformants which show impaired growth in both induced and non-induced media, thus showing a growth defect due to integrative disruption, are selected as well.

[0046] Transformants showing impaired growth are supposed to contain plasmids which produce anti-sense RNA to mRNAs critical for growth or survival. Growth is monitored by measuring growth-curves over a period of time in a device (Bioscreen Analyzer, Labsystems) which allows simultaneous measurement of growth-curves of 200 transformants.

[0047] Subsequently plasmids can be recovered from the transformants and the sequence of their inserts determined, thus revealing which mRNA they inhibit. In order to be able to recover the genomic or cDNA insert which has integrated into the *Candida* genome, genomic DNA is isolated, cut with an enzyme which cuts only once into the library vector (and estimated approx. every 4096 bp in the genome) and religated. PCR with primers flanking the insert will yield (partial) genomic or cDNA inserts as PCR fragments which can directly be sequenced. This PCR analysis (on ligation reaction) will also show us how many integrations occurred. Alternatively the ligation reaction is transformed to *E. coli* and PCR analysis is performed on colonies or on plasmid DNA derived thereof.

[0048] This method is employed for a genome wide search for novel *C. albicans* genes which are important for growth or survival.

Materials & Methods

Construction of pGal1PNiST-1

[0049] The backbone of the pGAL1PNiST-1 vector (integrative anti-sense *Sfi*-*Nol* vector) is pGEM11Zf(+) (Promega Inc.). First, the CaMAL2EcoRI/*Sai* promoter fragment from pDBV50 (D.H. Brown *et al.*) was ligated into *Eco*RI/*Sai*-opened pGEM11Zf(+) resulting in the intermediate construct pGEMMAL2P-1. Into the latter (*Msc*l/*CIP*) the CaURA3 selection marker was cloned as a *Eco*47III/*Xmn*I fragment derived from pRM2. The resulting pGEMMAL2P-2 vector was *Nol*II/*Hind*III opened in order to accept the *Nol*-stuffer-*Sfi* cassette from pPCK1NiSCYCT-1 (*Eag*l/*Hind*III fragment): pMAL2PNiST-1. Finally, the plasmid pGAL1PNiST-1 was constructed by exchanging the *Sai*II/*Ec136*II MAL2 promoter in pMAL2PNiST-1 by the *Xba*l/*Sma*1 GAL1 promoter fragment derived from pRM2GAL1P.

Construction of pGal1PSiST-1

[0050] The vector pGAL1PSiST-1 was created for cloning the small genomic DNA fragments (flanked by *S*_{fi} sites) behind the GAL1 promoter. The only difference with pGAL1PNiST-1 is that the hIFN β (stuffer fragment) insert fragment in pGAL1PSiST-1 is flanked by two *S*_{fi} sites in stead of a *S*_{fi} and a *N*_{ci} site as in pGAL1PNiST-1. To construct pGAL1PSiST-1 the *Eco*RI-*Hind*III fragment, containing hIFN β flanked by a *S*_{fi} and a *N*_{ci} site, of pMAL2pHiET-3 (unpublished) was exchanged by the *Eco*RI-*Hind*III fragment, containing hIFN β flanked by two *S*_{fi} sites, from YCp50S-S (an *E. coli* / *S. cerevisiae* shuttle vector derived from the plasmid YCp50, which is deposited in the ATCC collection (number 37419; Thrash *et al.*, 1985); an *Eco*RI-*Hind*III fragment, containing the gene hIFN β , which is flanked by two *S*_{fi} sites, was inserted in YCp50, creating YCp50S-S), resulting into plasmid pMAL2PSiST-1. The *ma2* promoter from pMAL2PSiST-1 (by a *Nae*I-*Fsp*I digest) was further replaced by the *ga1* promoter from pGAL1PNiST-1 (via a *Xba*I-*S*_{fi}I digest), creating the vector pGAL1PSiST-1.

10

Candida albicans genomic library

* Preparation of the genomic DNA fragments

15

[0051] A *Candida albicans* genomic DNA library with small DNA fragments (400 to 1,000 bp) was prepared. Genomic DNA of *Candida albicans* B2630 was isolated following a modified protocol of Blin and Stafford (1976). The quality of the isolated genomic DNA was checked by gel electrophoresis. Undigested DNA was located on the gel above the marker band of 26,282 bp. A little smear, caused by fragmentation of the DNA, was present. To obtain enrichment for genomic DNA fragments of the desired size, the genomic DNA was partially digested. Several restriction enzymes (*Alu*I, *Hae*III and *Rsa*I; all creating blunt ends) were tried out. The appropriate digest conditions have been determined by titration of the enzyme. Enrichment of small DNA fragments was obtained with 70 units of *Alu*I on 10 μ g of genomic DNA for 20 min. T4 DNA polymerase (Boehringer) and dNTPs (Boehringer) were added to polish the DNA ends. After extraction with phenolchloroform the digest was size-fractionated on an agarose gel. The genomic DNA fragments with a length of 500 to 1,250 bp were eluted from the gel by centrifugal filtration (Zhu *et al.*, 1985). *S*_{fi} adaptors (5' GTTGGCCTTTT) or (5' AGGCCAAC) were attached to the DNA ends (blunt) to facilitate cloning of the fragments into the vector. Therefore, a 8-mer and 11-mer oligonucleotide (comprising the *S*_{fi} site) were kinned and annealed. After ligation of these adaptors to the DNA fragments a second size-fractionation was performed on an agarose gel. The DNA fragments of 400 to 1150 bp were eluted from the gel by centrifugal filtration.

20

* Preparation of the pGAL1PSiST-1 vector fragment

25

[0052] The small genomic DNA fragments were cloned after the GAL1 promoter in the vector pGAL1PSiST-1. Qiagen-purified pGAL1PSiST-1 plasmid DNA was digested with *S*_{fi} and the largest vector fragment eluted from the gel by centrifugal filtration (Zhu *et al.*, 1985). Ligation with a control DNA fragment, flanked by *S*_{fi} sites, was performed as a control. The ligation mix was electroporated to MC1061 *E. coli* cells. Plasmid DNA of 24 clones was analyzed. In all cases the control fragment was inserted in the pGAL1PSiST-1 vector fragment.

30

* Upscaling

35

[0053] All genomic DNA fragments (450 ng) were ligated into the pGAL1PSiST-1 vector (20 ng). After electroporation at 2500V, 40 μ F circa 400,000 clones were obtained. These clones were pooled into three groups and stored as glycerol slants. Also Qiagen-purified DNA was prepared from these clones. A clone analysis showed an average insert length of 600 bp and a percentage of 91 for clones with an insert. The size of the library corresponds to 5 times the diploid genome. The genomic DNA inserts are sense or anti-sense orientated in the vector.

Candida albicans cDNA library

40

[0054] Total RNA was extracted from *Candida albicans* B2630 grown on respectively minimal (SD) and rich (YPD) medium as described by Chirgwin *et al.* in Sambrook *et al.* mRNA was prepared from total RNA using the Invitrogen Fast Track procedure.

50

[0055] First strand cDNA is synthesised with the Superscript Reverse Transcriptase (BRL) and with an oligo dT-*N*_{ci} Primer adapter. After second strand synthesis, cDNA is polished with Klenow enzyme and purified over a Sephadryl S-400 spun column. Phosphorylated *S*_{fi} adapters are then ligated to the cDNA, followed by digestion with the *N*_{ci} restriction enzyme. The *S*_{fi}/*N*_{ci} cDNA is then purified and sized on a Biogel column A150M.

[0056] First fraction contains approximately 38,720 clones by transformation, the second fraction only 1540 clones. Clone analysis:

Fr. I: 22/24 inserts, 16 \geq 1000 bp, 4 \geq 2000 bp, average size: 1500 bp.

55

Fr. II: 9/12 inserts, 3 \geq 1000 bp, average size: 960 bp cDNA was ligated in a *N*_{ci}/*S*_{fi} opened pGAL1PNiST-1 vector (anti-sense)

Candida transformation

[0057] The host strain used for transformation is a *C. albicans* ura3 mutant, CAI-4, which contains a deletion in orotidine-5'-phosphate decarboxylase and was obtained from William Fonzi, Georgetown University (Fonzi and Irwin). CAI-4 was transformed with the above described cDNA library or genomic library using the Pichia spheroplast module (Invitrogen). Resulting transformants were plated on minimal medium supplemented with glucose (SD, 0.67% or 1.34% Yeast Nitrogen base w/o amino acids + 2% glucose) plates and incubated for 2-3 days at 30°C.

Screening for mutants

[0058] Starter cultures were set up by inoculating each colony in 1 ml SD medium and incubating overnight at 30°C and 300 rpm. Cell densities were determined using a Coulter counter (Coulter Z1; Coulter electronics limited). 250.000 cells/ml were inoculated in 1 ml SD medium and cultures were incubated for 24 hours at 30°C and 300 rpm. Cultures were washed in minimal medium without glucose (S) and the pellet resuspended in 650 µl S medium. 8 µl of this culture is used for inoculating 400 µl cultures in a Honeywell-100 plate (Bioscreen analyzer; Labsystems). Each transformant was grown during three days in S medium containing LiAc; pH 6.0, with 2% glucose/2% maltose or 2% galactose/2% maltose respectively while shaking every 3 minutes for 20 seconds. Optical densities were measured every hour during three consecutive days and growth curves were generated (Bioscreen analyzer; Labsystems).

[0059] Growth curves of transformants grown in respectively anti-sense non-inducing (glucose/maltose) and inducing (galactose/maltose) medium are compared and those transformants showing impaired growth upon anti-sense induction are selected for further analysis. Transformants showing impaired growth by virtue of integration into a critical gene are also selected.

Isolation of genomic or cDNA inserts

[0060] Putatively interesting transformants are grown in 1.5 ml SD overnight and genomic DNA is isolated using the Nucleon MI Yeast kit (Clontech). Concentration of genomic DNA is estimated by analyzing a sample on an agarose gel.

[0061] 20 ng of genomic DNA is digested for three hours with an enzyme that cuts uniquely in the library vector (SacI for the genomic library; PstI for the cDNA library) and treated with RNase. Samples are phenol/chloroform extracted and precipitated using NaOAc/ethanol.

[0062] The resulting pellet is resuspended in 500 µl ligation mixture (1 x ligation buffer and 4 units of T4 DNA ligase; both from Boehringer) and incubated overnight at 16°C.

[0063] After denaturation (20 min 65°C), purification (phenol/chloroform extraction) and precipitation (NaOAc/ethanol) the pellet is resuspended in 10 µl MilliQ (Millipore) water.

PCR analysis

[0064] Inverse PCR is performed on 1 µl of the precipitated ligation reaction using library vector specific primers (oligo23 5' TGC-AGC-TCG-ACC-TCG-ACT-G 3' and oligo25 5' GCG-TGA-ATG-TAA-GCG-TGA-C 3' for the genomic library; 3pGALNistPCR primer :5TGAGCAGCTGCCGTCGCGC 3' and 5pGALNistPCR primer: 5'GAGTTATACCTGCAGCTCGAC 3' for the cDNA library; both from Eurogentec) for 30 cycles each consisting of (a) 1 min at 95 °C, (b) 1 min at 57 °C, and (c) 3 min at 72 °C. In the reaction mixture 2.5 units of Taq polymerase (Boehringer) with TaqStart antibody (Clontech) (1:1) were used, and the final concentrations were 0.2 µM of each primer, 3 mM MgCl2 (Perkin Elmer Cetus) and 200 µM dNTPs (Perkin Elmer Cetus). PCR was performed in a Robocycler (Stratagene).

Sequence determination

[0065] Resulting PCR products were purified using PCR purification kit (Qiagen) and were quantified by comparison of band intensity on EtBr stained agarose gel with the intensity of DNA marker bands. The amount of PCR product (expressed in ng) used in the sequencing reaction is calculated as the length of the PCR product in basepairs divided by 10. Sequencing reactions were performed using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit according to the instructions of the manufacturer (PE Applied Biosystems, Foster City, CA) except for the following modifications.

[0066] The total reaction volume was reduced to 15 µl. Reaction volume of individual reagents were changed accordingly. 6.0 µl Terminator Ready Reaction Mix was replaced by a mixture of 3.0 µl Terminator Ready Reaction Mix + 3.0 µl Half Term (GENPAK Limited, Brighton, UK). After cycle sequencing, reaction mixtures were purified over Sephadex G50 columns prepared on Multiscreen HV opaque microtiter plates (Millipore, Molsheim, Fr) and were dried in a speedVac. Reaction products were resuspended in 3 µl loading buffer. Following denaturation for 2 min at 95°C, 1 µl of sample was applied on a 5% Long Ranger Gel (36 cm well-to-read) prepared from Singel Packs according to the supplier's instructions (FMC BioProducts, Rockland, ME). Samples were run for 7 hours 2X run on a ABI 377XL DNA sequencer. Data collection version 2.0 and Sequence analysis version 3.0 (for basecalling) software packages are from PE Applied Biosystems. Resulting sequence text files were copied onto a server for further analysis.

Sequenc analysis

[0067] Nucleotide sequences were imported in the VectorNTI software package (InforMax Inc, North Bethesda, MD, USA), and the vector and insert regions of the sequences were identified. Sequence similarity searches against public and commercial sequence databases were performed with the BLAST software package (Altschul et al., 1990) version 1.4. Both the original nucleotide sequence and the six-frame conceptual translations of the insert region were used as query sequences. The used public databases were the EMBL nucleotide sequence database (Stoesser et al., 1998), the SWISS-PROT protein sequence database and its supplement TrEMBL (Bairoch and Apweiler, 1998), and the ALCES *Candida albicans* sequence database (Stanford University, University of Minnesota). The commercial sequence databases used were the LifeSeq® human and PathoSeq™ microbial genomic databases (Incyte Pharmaceuticals Inc., Palo Alto, CA, USA), and the GENESSEQ patent sequence database (Derwent, London, UK). Three major results were obtained on the basis of the sequence similarity searches: function, novelty, and specificity. A putative function was deduced on the basis of the similarity with sequences with a known function, the novelty was based on the absence or presence of the sequences in public databases, and the specificity was based on the similarity with vertebrate homologues.

Methods

[0068] Blastx of the nucleic acid sequences against the appropriate protein databases: Swiss-Prot for clones of which the complete sequence is present in the public domain, and paorfp (PathoSeq™) for clones of which the complete sequences is not present in the public domain.

[0069] The protein to which the translated nucleic acid sequence corresponds to is used as a starting point. The differences between this protein and our translated nucleic acid sequences are marked with a double line and annotated above the protein sequence. The following symbols are used: a one-letter amino acid code or the ambiguity

code X is used if our translated nucleic acid sequence has another amino acid on a certain position,

the stop codon sign * is used if our translated nucleic acid sequence has a stop codon on a certain position,

The letters fs (frame shift) are used if a frame shift occurs in our translated nucleic acid sequence, and another reading frame is used,

the words ambiguity or ambiguities are used if a part of our translated nucleic acid sequence is present in the proteins, but not visible in the alignments of the blast results,

The phrase missing sequence is used if the translated nucleic acid sequence does not comprise that part of the protein.

Blastx: compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

Screening for compounds modulating expression of polypeptides critical for growth and survival of *C. albicans*

[0070] The method proposed is based on observations (Sandbaken et al., 1990; Hinnebusch and Liebman 1991; Ribogene PCT WO 95/11969, 1995) suggesting that underexpression or overexpression of any component of a process (e.g. translation) could lead to altered sensitivity to an inhibitor of a relevant step in that process. Such an inhibitor should be more potent against a cell limited by a deficiency in the macromolecule catalyzing that step and/or less potent against a cell containing an excess of that macromolecule, as compared to the wild type (WT) cell.

[0071] Mutant yeast strains, for example, have shown that some steps of translation are sensitive to the stoichiometry of macromolecules involved. (Sandbaken et al.). Such strains are more sensitive to compounds which specifically perturb translation (by acting on a component that participates in translation) but are equally sensitive to compounds with other mechanisms of action.

[0072] This method thus not only provides a means to identify whether a test compound perturbs a certain process but also an indication of the site at which it exerts its effect. The component which is present in altered form or amount in a cell whose growth is affected by a test compound is potentially the site of action of the test compound.

[0073] The assay to be set up involves measurement of growth of an isogenic strain which has been modified only in a certain specific allele, relative to a wild type (WT) *C. albicans* strain, in the presence of R-compounds. Strains can be ones in which the expression of a specific essential protein is impaired upon induction of anti-sense or strains which carry disruptions in an essential gene. An in silico approach to finding novel essential genes in *C. albicans* will be performed. A number of essential genes identified in this way will be disrupted (in one allele) and the resulting strains can be used for comparative growth screening.

Assay for High Thr ughput scr ening f r drugs

[0074] 35 µl minimal medium (S medium + 2% galactose + 2% maltose) is transferred in a transparent flat-

EP 0 982 401 A2

bottomed 96 well plate using an automated pipetting system (Multidrop, Labsystems). A 96-channel pipettor (Hydra, Robbins Scientific) transfers 2.5 μ l of R-compound at 10^{-3} M in DMSO from a stock plate into the assay plate.

[0075] The selected *C. albicans* strains (mutant and parent (CAI-4) strain) are stored as glycerol stocks (15%) at -70°C. The strains are streaked out on selective plates (SD medium) and incubated for two days at 30°C. For the parent strain, CAI-4, the medium is always supplemented with 20 μ g/ml uridine. A single colony is scooped up and resuspended in 1 ml minimal medium (S medium + 2% galactose + 2% maltose). Cells are incubated at 30°C for 8 hours while shaking at 250 rpm. A 10 ml culture is inoculated at 250,000 cells/ml. Cultures are incubated at 30°C for 24 hours while shaking at 250 rpm. Cells are counted in Coulter counter and the final culture (S medium + 2% galactose + 2% maltose) is inoculated at 20,000 to 50,000 cells/ml. Cultures are grown at 30°C while shaking at 250 rpm until a final OD of 0.24 (+/- 0.04) 6nM is reached.

[0076] 200 μ l of this yeast suspension is added to all wells of MW96 plates containing R-compounds in a 450 μ l total volume. MW96 plates are incubated (static) at 30°C for 48 hours.

[0077] Optical densities are measured after 48 hours.

[0078] Test growth is expressed as a percentage of positive control growth for both mutant (x) and wild type (y) strains. The ratio (x/y) of these derived variables is calculated.

15 References

[0079]

Thrash C., Bankier A. T., Barrell B. G. and Sternblitz R. (1985) *Proc. Natl. Acad. Sci. USA* **82**: 4374-4378.

20 Blin N. and Stafford D. W. (1976) *Nucleic Acids Res.* **3**: 2303-2308.

Zhu J., Kempenaers W., Van Der Straeten D., Contreras R., and Fiers W. (1985) *Bio/Technology* **3**: 1014-1016.

25 Sambrook J., Fritsch E.F. and Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Fonzi W. and Irwin H. (1993) *Genetics* **134**:717-728. Altschul S.F, Gish W., Miller W., Myers E.W., Lipman D.J. (1990) *J. Mol. Biol.* **215**(3):403-410.

30 Bairoch A. and Apweiler R. (1998) *Nucleic Acids Res.* **26**(1):38-42.

Stoesser G., Moseley M.A., Sleep J., McGowran M., Garcia-Pastor M., Sterk P. (1998) *Nucleic Acids Res.* **26**(1):8-15.

Chien et al., (1991) *Proc. Natl. Acad. Sci USA* **88**, 9578-9582.

35 Sandbaken M.G. Lupisella J.A., DiDomenico B. and Chakraburty K., *J. Biol. Chem.* **265**:15838-15844, 1990.

Hinnebusch A.G. and Liebman S.W., in : *The Molecular Biology of the Yeast Saccharomyces*, Broach J.R., Pringle J.R. and Jones E.W., eds., CSH Laboratory Press; NY 1991.

40 Patent application RiboGene Inc., PCT WO 95/11969, 1995.

Brown, D.H. et al., *Molecular General Genetics* **251** (1) 75-80, 1996.

45

50

55

EP 0 982 401 A2

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5

(i) APPLICANT:
 (A) NAME: Janssen Pharmaceutica
 (B) STREET: Turnhoutseweg 30
 (C) CITY: Beerse
 (E) COUNTRY: Belgium
 (F) POSTAL CODE (ZIP): B-2340
 (G) TELEPHONE: +32 (0)14/60.21.11
 (H) TELEFAX: +32 (0) 14/60.28.41

10

(ii) TITLE OF INVENTION: DRUG TARGETS IN CANDIDA ALBICANS

15

(iii) NUMBER OF SEQUENCES: 72

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: GB 9817796.7
 (B) FILING DATE: 14-AUG-1998

20

(2) INFORMATION FOR SEQ ID NO: 1:

25

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 255 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AACGTTGGTCAAAAGGCTA TACTGGTGTATCCACCGCAG ATGAAGAGCA AGTTAAATCA	60
ACTCTTTGTC ATTAATGCT GTACTTGTT TCATTTTATT TGCTGGCATT TAAAGAATAC	120
CCATAGTTCA GAAAATAAAA TTGAAAAATT TAAAAAAA CGCAATATCA TTCACTTTTT	180
TTGTTTTTTT GACAATAATA TTAATATGTA GTTACCAATG TTTTAGATT TTATATGTTT	240
TGAAAAAATA GTTTG	255

35

(2) INFORMATION FOR SEQ ID NO: 2:

40

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 648 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

50

55

EP 0 982 401 A2

	AACCTCTTAT TCGGTTCTAG TGTCTCAATT GGTTATCCAT TAACATCTAT TCCCAACTCC	60
5	ATCATTATTG GCAATAAATA AATGGGTGTT ATATCTATTG GTAATAACTA AACTGGTGT	120
	AATTCAATTG CAATATGGTC ATGACATTG AAAGTGTAC TGTCTGGTT TACATATTCT	180
	ACAGGTTACA ACTATTGATT GGTTAGAAGT TTGGTTCAA CATCACCTGT TGCTAAGAAT	240
	AAATGTTGGT CATATCAATT GAATCATTG TTGGTGTAT GGTAAGTAAA TGCTGGTTAT	300
10	ATCTATTATC TACAACCACC AAGTGATAAA TGCTGAACCG TAGTCACCAA CTGTTATGCT	360
	GGTTGTATCT ATTGACTAA ACTACCCCTAG GGATAAAATGC TGAAACCGTGG TTACAACTG	420
	TTATGCTGGT TGTATCTATT AACTGCAACC ACCAAATGAT AAATGCTGAA CCATAATTAC	480
	CAACTGTTAC ATTGCTGGTA CTACATTAAG AATAAATGCT GCATCTACAA GTACCACCTG	540
15	TTGTGTTAAT AAATGCTGCA CCTGCTAGTA CAACTGTTGC TGGTCATGAT AGTTACTACA	600
	CATTACACAC CAGACAGTGG CAAACAAAGGT TATGTAGAAA CCAACGTT	648

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 904 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25 (iii) HYPOTHETICAL: NO

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:	
30	AACTGCTTG TGAAGACGAA CATCACAAACC ACAATCATGG TCATAACCAA AATCACAAATC	60
	ATGTTGCTCC TATTCTTACA ACAGCTGGAC AATCATTAAA TAATAAAATT GATACATCTA	120
	AAGTACAGC TCTCAACATG GCCAACTCTG CTGACGATCT AGCAAAAGTT TTCAAAGATT	180
35	CGACTAAAAA ATATCAAATC AAACCAATTAA TCAAATCAGA CAGTGATGAA CAAATGATTA	240
	TCAACATTCC ATTCTTAAT GGTAGTGTCA AATTGTATTG GATAATTCTA CGTACCAATG	300
	GGGATTGTA TTGTCACAA ACAATAAAAT TATTCAAAAA TGACACATCA ATTGATTTG	360
	ATAATGTGGA TTCAAGAAA CCAATACAGG TGTTAACTCA TCCTCAAGTT GGTGTTGCTA	420
40	ATAATGATAG CGATGATCTT CCAGAGTTT TGGAAATCAA TAACGATGAC GATTTGTCG	480
	AACATTATGT GTCTCGACAT AAATTCACTG GGGTAAATCA ATTGACAATA TTTATTGAAG	540
	ATATTATGAA TGAAGGAGAA GAAGAGTGTGTC ATTACATTG AATTGATTTG AGAGGGAAAT	600
45	TCACTGAATT AAACAAAGAC CCTGTCATTA CATTATATGA ACTGGCTGCT AATCCTGCTG	660
	ATCATAAGAA TTTAACGATT GTTGAAGATC AAAATCTAGC ATAAAACAAA GAAGTGAAG	720
	GTATCAGATA AGCTGGTTAC ATTACAATTG ATCTAATTAA GAATCTCAAG GTATTTAAAT	780
	TTGCCGTTT GCGATAATAT AACATGGTCA AGAACGTTGA ATCGATTACG TTAATGGTTT	840
50	AGCTAATTGA TTTTGTAGGAT CGAGTATTAA GAGTGAATAA ACAATAAACAA AGAATGATGA	900
	ATTG	904

EP 0 982 401 A2

(2) INFORMATION FOR SEQ ID NO: 4:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 232 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Ser Cys Glu Asp Glu His His Asn His Asn His Gly His Asn Gln Asn
1 5 10 15

15

His Asn His Val Ala Pro Ile Pro Thr Thr Ala Gly Gln Ser Leu Asn
20 25 30

20

Asn Lys Ile Asp Thr Ser Lys Val Thr Ala Leu Asn Met Ala Asn Ser
35 40 45

25

Ala Asp Asp Leu Ala Lys Val Phe Lys Asp Ser Thr Lys Lys Tyr Gln
50 55 60

30

Ile Lys Pro Ile Ile Lys Ser Asp Ser Asp Glu Gln Met Ile Ile Asn
65 70 75 80

Ile Pro Phe Leu Asn Gly Ser Val Lys Leu Tyr Ser Ile Ile Leu Arg
85 90 95

Thr Asn Gly Asp Leu Tyr Cys Pro Lys Thr Ile Lys Leu Phe Lys Asn
100 105 110

Asp Thr Ser Ile Asp Phe Asp Asn Val Asp Ser Lys Pro Ile Gln
115 120 125

Val Leu Thr His Pro Gln Val Gly Val Ala Asn Asn Asp Ser Asp Asp
130 135 140

Leu Pro Glu Phe Leu Glu Ser Asn Asn Asp Asp Asp Phe Val Glu His
145 150 155 160

35

Tyr Val Ser Arg His Lys Phe Thr Gly Val Asn Gln Leu Thr Ile Phe
165 170 175

Ile Glu Asp Ile Tyr Asp Glu Gly Glu Glu Cys His Leu His Ser
180 185 190

40

Ile Glu Leu Arg Gly Glu Phe Thr Glu Leu Asn Lys Asp Pro Val Ile
195 200 205

Thr Leu Tyr Glu Ser Ala Ala Asn Pro Ala Asp His Lys Asn Leu Thr
210 215 220

45

(2) INFORMATION FOR SEQ ID NO: 5:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 608 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

55

(iii) HYPOTHETICAL: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

10

AACCTACAAA	AGACTCACAT	GTGCTGTACA	ATAAATTCT	GGATAAGCAT	ATAAGTGATG	60
AGCAACTATC	ACACTTACTC	GACAATCATA	AACCCAATCT	AGTGACTACC	ACAACCTTAA	120
TTGATTCTAT	CAAAGAAAGT	GAACGTGTTAT	ATAATACCAT	GGACAGTTTG	ATGATAAAAAT	180
CCATCAATT	TCCTGCAGCC	ATGTACCGT	CAAATGACAA	CAATTACAA	TCACCAATCG	240
AGTATTATC	TAACAGAGTA	AAATTGCTCA	CACAAGAGTT	ATACGAAAGAT	TCAGTCAAAT	300
ATGGCAAGTT	TCTACAGAGT	GGTAATAATC	ATATATATCA	ATTACGAAAGT	AGGATTTAC	360
AGACCTTGA	TCAGTTGTCA	GAGAGTCACT	ATTCTTTAAA	TGAACATATAT	AATAAAGACA	420
TGTCTTACGC	AGAACACATTA	CACGGATCTT	TCAAGAAATG	GGATCAACAA	AGAAATAAAG	480
TATTGTCCAA	AGTGAATCT	ATAAAAAGTG	ATACAAGCAA	ACATGGAGCC	AAATTATTCA	540
CCTTATTAGA	TGAAGTTAAT	GATGTTGATG	ACGAGATCAA	ACTTTGGAA	GCAAAACTAC	600
AGCAGGTT						608

20

(2) INFORMATION FOR SEQ ID NO: 6:

25

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1497 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(iii) HYPOTHETICAL: NO

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

40

45

50

GATATCTGCA	GAATTCGGCT	TCTCTCTCAT	CTTCACACAA	TGCATTTAC	AAGTAGCCTA	60
CTAGCCACCT	TGATATGGTT	TACATTACCG	GTTCAAAGTT	TGAATACTGA	ATCTAGGACA	120
ACTTCARATA	ACACAATATC	AATACTTACA	AACCATTTC	AAATACTAAA	GGATTGCTA	180
CCATATAGCA	AACTTCTAA	ACCGCAAATC	AAGGAATCCA	GACCGTTGAT	TAAGTTCTG	240
AGAGATGGAG	TGCCAATAAA	TTTCCACAGG	GCTCCGGCTA	TAATAATGAA	ATCGAACAAA	300
ACAGACGATT	TAGTCAGGAA	TAGCAATAAA	ACAATGGTGC	TAACTGAAAT	AAAAACGATT	360
ACTGAATTG	CAACTACCAC	TGTTCCCT	ACACAAGAAAT	TTCAAGCACT	ACAGATAAAC	420
CTTAACACGT	TATCAATAGA	GACTTCACAA	CCAACATTCC	AATCCCATGA	CTTCCACCG	480
ATTACCAATTG	AAGACACACC	CAAAACACTA	GAACCGAGAG	AATCGTCAGA	TGCTTTGCAG	540
AGGGATGCAT	TTGATCAAAT	TAAGAAACTA	AAAAAATTGG	TATTGGATT	GAGACTTGAA	600
ATGAAAGAGC	AAACAAAGAG	TTCAACGAT	CAATTAGTGG	ATATATATAC	CGCAAGAAGT	660
ATTGTTCCAA	TTTATACTAC	ACATATCGTC	ACTTCGGCGA	TTCCATCGTA	TGTACCAAAA	720
GAAGAAGTAA	TGGTTTCACA	TGATACTGCA	CCAATTGTAA	GTCGTCTAG	ACAGATATT	780

55

EP 0 982 401 A2

5	CCAGTATCTC AACGAATTGA TACTATCTCA AACATAAAA TGAATGGAAA AAATATATTG	840
	AACRACAATC CTCCGCCAA TTCAGTTTA ATAGTTCTC AGTTCACTT CCATGAAAGA	900
	ATGCCACCA AAACCGAAGT AGCTTATATG AAACCAAAA TTGCTGGAC CAACTTCCA	960
	ACCACTACTG CAACGTCAT GTTGACAAT TTTATTTAA AAAATCTGT TGACGAAACG	1020
	GATTCTGAAA TTGATAGTGG TGAAACTGAA TTGCTGACG ATTATTATTA CTATTATAGT	1080
10	TACGAGATG ATGGTAAAGA AGACGATAGT GATGAGATTA CGGCTCAAAT ACTATTATCC	1140
	AATTCAAAAT TAGGCACGAA GACGCCAAAT TTTGAGGATC CTTTGAAACA AATCAATT	1200
	GAAGACAATA AAGTAATATC TGTTAATACA CCAAAGACAA AGAACCTAC TACAACAGTA	1260
	TTTGGCACTT CTACTAGTGC ATTATCACT TTTGAAAGTA CAATATTGA AATTCCAAA	1320
15	TTCTTTATG GTAGCAGAAG AAAACAATG AGCTCATTCA AATAAGAA CAGTACAATC	1380
	AAATTGATG TGGTGTGATTG GATATTGAA AGTGGTACTA CCAATGAGAA AGTACATGGA	1440
	TTAGTGTGG TGTCTAGTGG TGTTCTACTA GGAACCTGTC TATTGTCAT TTTGTAG	1497

(2) INFORMATION FOR SEQ ID NO: 7:

20	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 485 amino acids
	(B) TYPE: amino acid
	(C) STRANDEDNESS:
	(D) TOPOLOGY: unknown
25	(ii) MOLECULE TYPE: peptide
	(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

30	Met His Phe Thr Ser Ser Leu Leu Ala Thr Leu Ile Trp Phe Thr Leu	
	1 5 10 15	
	Pro Val Gln Ser Leu Asn Thr Glu Ser Arg Thr Thr Ser Asn Asn Thr	
	20 25 30	
35	Ile Ser Ile Leu Thr Asn His Phe Gln Ile Leu Lys Asp Leu Leu Pro	
	35 40 45	
	Tyr Ser Lys Thr Ser Lys Pro Gln Ile Lys Glu Ser Arg Pro Leu Ile	
	50 55 60	
	Lys Val Ser Arg Asp Gly Val Pro Ile Asn Phe His Arg Ala Pro Ala	
40	65 70 75 80	
	Ile Ile Met Lys Ser Asn Lys Thr Asp Asp Leu Val Arg Asn Ser Asn	
	85 90 95	
	Lys Thr Met Val Leu Thr Glu Ile Lys Thr Ile Thr Glu Phe Ala Thr	
	100 105 110	
45	Thr Thr Val Ser Pro Thr Gln Glu Phe Gln Ala Leu Gln Ile Asn Leu	
	115 120 125	
	Asn Thr Leu Ser Ile Glu Thr Ser Thr Pro Thr Phe Gln Ser His Asp	
	130 135 140	
50	Phe Pro Pro Ile Thr Ile Glu Asp Thr Pro Lys Thr Leu Glu Pro Glu	
	145 150 155 160	
	Glu Ser Ser Asp Ala Leu Gln Arg Asp Ala Phe Asp Gln Ile Lys Lys	

EP 0 982 401 A2

	165	170	175
	Leu Glu Lys Leu Val Leu Asp Leu Arg	Leu Glu Met Lys Glu Gln Gln	
	180	185	190
5	Lys Ser Phe Asn Asp Gln Leu Val Asp Ile Tyr	Thr Ala Arg Ser Ile	
	195	200	205
	Val Pro Ile Tyr Thr Thr His Ile Val Thr Ser Ala Ile Pro Ser Tyr		
	210	215	220
10	Val Pro Lys Glu Glu Val Met Val Ser His Asp Thr Ala Pro Ile Val		
	225	230	235
	Ser Arg Pro Arg Thr Asp Ile Pro Val Ser Gln Arg Ile Asp Thr Ile		
	245	250	255
	Ser Lys His Lys Met Asn Gly Lys Asn Ile Leu Asn Asn Asn Pro Pro		
	260	265	270
15	Pro Asn Ser Val Leu Ile Val Pro Gln Phe Gln His Glu Arg Met		
	275	280	285
	Ala Thr Lys Thr Glu Val Ala Tyr Met Lys Pro Lys Ile Val Trp Thr		
	290	295	300
20	Asn Phe Pro Thr Thr Ala Thr Ser Met Phe Asp Asn Phe Ile Leu		
	305	310	315
	Lys Asn Leu Val Asp Glu Thr Asp Ser Glu Ile Asp Ser Gly Glu Thr		
	325	330	335
25	Glu Leu Ser Asp Asp Tyr Tyr Tyr Ser Tyr Glu Asp Asp Gly		
	340	345	350
	Lys Glu Asp Asp Ser Asp Glu Ile Thr Ala Gln Ile Leu Leu Ser Asn		
	355	360	365
30	Ser Glu Leu Gly Thr Lys Thr Pro Asn Phe Glu Asp Pro Phe Glu Gln		
	370	375	380
	Ile Asn Ile Glu Asp Asn Lys Val Ile Ser Val Asn Thr Pro Lys Thr		
	385	390	395
	Lys Lys Pro Thr Thr Val Phe Gly Thr Ser Thr Ser Ala Leu Ser		
	405	410	415
35	Thr Phe Glu Ser Thr Ile Phe Glu Ile Pro Lys Phe Phe Tyr Gly Ser		
	420	425	430
	Arg Arg Lys Gln Ser Ser Ser Phe Lys Asn Lys Asn Ser Thr Ile Lys		
	435	440	445
40	Phe Asp Val Phe Asp Trp Ile Phe Glu Ser Gly Thr Thr Asn Glu Lys		
	450	455	460
	Val His Gly Leu Val Leu Val Ser Ser Gly Val Leu Leu Gly Thr Cys		
	465	470	475
	Leu Leu Phe Ile Leu		
45		485	

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1651 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

	GAGCTCTTCC AGAGGAAACA AGCGGAAGAA GCACAAACGAA AGAAGGAATT TGACCAAAAG	60
	GCCGAAATTCA TCAAAGCATC ATTACTGAA ATGCGCCGAA GAGAAATAGA GAGGCGGAAA	120
10	CAGCAAAAGG AAAGGAAACA AAGACAAAG GAGCACGGAAG CAAAGAGGGA TATCAGGATA	180
	CAACAACTTT CAGAGCAGGA TTCACGGAGT AATCAAACTA AAGAAGAAGA GGAAGTGTTC	240
	AAGAAGGCC GGTCTACTAA TTCCGGAGCA GACGAGACTG GTTGATGTC AGATAAAGAG	300
	TTTGATGATT CTGCATATTG ACCCGATTAT TTGTTGAAAG AGAATTGAG GATAAAACCA	360
15	AATCATCCAG ATACAAATCA TAAAACAAA AAATATACTG AGAATGTGGT TGAAAATCTA	420
	GATTCTCCAC CAAATGATAC ATCTGCGTAC AATTCAAGT TTCAATGATGA AACTAATATT	480
	CAAAATGAGA TCCAAATACC AGAAAATGAC GAGTATGTAC CACAGATGAA AGCTACATCC	540
20	AGTGTCAATA ATACCACCAT CCCTGCACAA AGAAGACATG AGTCACCTTC CACTTCTGAA	600
	AACAAAAGAA GGAATTTGA AACAGCCGAC GTTGGGTTG ATGGGTTAGA TTCCCCAGTG	660
	CGGGCACAAC CAGAAATATC TGGAAAATCC AAGTCTCCGA TAATCCCTGA TGTAACTATT	720
	TTACTGGACG AAGAGACTGA AACTCCTGAA GCAAATGCTG TGCAGGACAA TAGTACATAT	780
25	ATTCCCTCAGG GGTCTTTAGG ACACGAATT AGAAATATT TGGAAAGAGCA TCCACGTCAA	840
	GTAAAGAATA AACAAAATTC TGGTGTGCT TTTGCATTTC CGAATGCTTC CAAGAATACC	900
	GAAAACAAAC TCCACTCTAA TTCAAAGAT AAAGATGAAG GAATAATTGA TGTGAAAGCT	960
30	TACGTACCTG ATGTCAAAGC AGCAACTTCA AACACCACCC CAGCAACAGG ACAAAACATCA	1020
	GCAAGGTCGG AAAAATGCC ACCCTTACCT ACTCATATTG CAAATCCATC GACCATGAAT	1080
	GAAGCTCGAC CTCATCCAAC AACTCCACAT AAAAGATCAA AAGTCATTTT CGATTAAAAA	1140
	GATTAGAAC AAAAGTTAGG TAATGATATT GAGGATTTGG ATTTTAAGGA TATGTATGAG	1200
35	AGTTGCCCTG ACCATTCAAG TAAGGAAACA CCTAAAGACG ATATTTAAC CGGTTCTAAA	1260
	AGAAGACTTT ATACATATAAC CGATGGAACA TCAAAGGCTG AAACGTTATC TACACCAATG	1320
	AACAAAATC CTGTTCGTGG ACATAGTACC AAGAAAAAGC TTAGTATGTT GGACATGCAT	1380
40	GGCTCTTCTA AAATTCAAAG TCTTTTACCT CCACAAACCC CACAAATGTC AATTGATCCT	1440
	TCTGTTCCA AGCAAGTGTG GGCTAAATAC GTTGATGCAA TCTTGACTTA TCAAAGAGAA	1500
	TTTTCAATT ATAAAAAAAGT GATTGTTCAA TACCAAATGG AACGGATAAA CAAAGACCTT	1560
	GAACATTTG ACCATATAAA TGATGGTCA CACACTGAGA ATTTGGATAC TTTCAAGCAT	1620
45	TGTTTAGAAC AAGATTATTG GTTAGTTGA C	1651

(2) INFORMATION FOR SEQ ID NO: 9:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 463 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AACCTGTTGA CGCGTTGTCT	TTTCTACCC CACGTTAAC	AATCTGCCA GTCAATTAC	60
TAGCCAAATA AACTTGTAGAC	TCACAACTCT AACACTGACT	CGCCCCCCCC TGTTAACT	120
10 CTAATTACT TCACAGAGCC	TTTACTACCT TAATTTAAGA	TTATCTATTG TTTCTGTTCT	180
TTTGCATCA CCCTGACTCG	TTTTTTTTC AGCCAGTTT	TTCGTAAAT CTGACCAAAA	240
ATTTACACT CTAATTAAA	ACTCTAAATA ACAATTAAA	CTCAATTCA ACAGTCCTT	300
15 CTGCTCATTC TGACTCTCT	CTATTGTCTT TTGACTTTT	GTGIGTGACT ATTTTCATGA	360
TCACCCCGTT TCTTGCATTT	TTTCAGTCA ACTTTTCTC	AAAATCAAGC CAAAAAAACA	420
CATTAACTG CCTATACAAAC	GCAAACCTAT TCAAAACAAG	GTT	463

(2) INFORMATION FOR SEQ ID NO: 10:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 582 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25 (iii) HYPOTHETICAL: NO

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AACCTCCCCG TAAACCACTT	CTAGGTATAC CATTTCATCT	GACTGAATAA CTGGTTAGTC	60
GATTGTTGT TGAAGAAAAG	TGACCACTA GTTTTTCTG	CCAACATTT TTGCGATGAG	120
CCGTCGACGC GTTGTCTTT	TCTACCCAC GTTTAACAAAT	CTTGCAGTC AATTCCCTAG	180
35 CCAAATAAAC TTAGACTCA	CAACTCTAAC ACTGACTCGT	GCCCCCTGT TAAACTCTA	240
AATTACTCA CAGAGCCTTT	ACTACCTAA TTTAAGATTA	TCTATTGTTT CTGTTTTTT	300
GCAATCACCC TGACTCGTT	TTTTTCAGC CAGTTTTTC	GTAAAATCTG ACCAAAAATT	360
40 TACAACCTCA ATTAAAAT	CTAAATAACA ATTTAAACTC	AATTCAACACA AGTCCTCTG	420
CTCATTCTGA GTCTTCTCTA	TTGTCTTTG ACTTTTGTG	TGTGACTATT TTCATGATCA	480
CCCCGTTCT TGCATTTTT	TCAGTCAACT TTTCTCAAA	ATCAAGCCAA AAAAACACAC	540
CTTTAACTAC CTATACAAACG	CAAACCTATT CAAACAAAGG	TT	582

45 (2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1066 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 183
 (D) OTHER INFORMATION:/note= "W = A or T"

5 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 564
 (D) OTHER INFORMATION:/note= "Y = C or T"

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AACCATAAAAT	ATGCCAAGAT	TTAACACAAGT	TGATGTATTTC	ACCAATGTCA	AATATTTGGG	60
TAATCCAGTT	GCCGTTATTT	ATGATAGTGA	TAATTTAAC	ACTCAAGAAA	TGCXXXXXAT	120
TGCTCGATGG	ACAAATTTAT	CAGAAACAAAC	ATTTATATTG	ACTCCAAAAT	CATCAATTGC	180
TGWTTATAGT	ATTAGAATTT	TCACTTCTGG	TGGGAATGAA	TTACCATTTG	CTGGTCATCC	240
TACTTTAGGT	ACTGCATTG	CATTATTGGA	AGATGGTAAA	ATAAAACCAA	ATGACAATGG	300
ACAAATAATT	CAAGAATGTG	GTGCTGGATT	AGTGAAAATA	TCCGTTGAAA	AAACACCTAA	360
TAATAATAGT	AATGAGTTGC	CGTTTTGTT	ATCTTTGAA	TTACCATATT	TCAAATTTCA	420
TGAAATTGAT	GACAAAGTAA	TCGAGGAATT	ACAACATTCA	TGGAATGGAA	CCAATATTAT	480
TGGTAAACCG	GTACTTATTG	ATGCTGGTCC	AAAATGGCA	GTTTCCAAC	TTGGCTCCGG	540
TAAAGAAGTA	TTAGACTTGA	ATGYTGATT	AGCACAAATT	GAGAGATTAA	GTTTAGAAAA	600
TGGTTGGACA	GGAATTGGTG	TCTTGGAAA	ACATAATGAA	AATGGTGATT	CGGTCGAATT	660
GAGAAATATT	GCTCTGCTG	TTGGAGTCGC	TGAAGATCCT	GCTTGTGGAA	GTGGATCAGG	720
TGCTATTGGA	GCATATTGG	CAAATCACGT	TTCAATGAA	AAGGAAAAT	TTACAATTGA	780
TATTTCTCAA	GGTAAACCAA	TTGAAAGAGA	TGCTAAGATT	CAAGTTAAAG	TTAATCGTCT	840
TACCAACAA	AATGGTGATT	TATCTATTCA	TGTTGGTGGT	CATGCCATCA	CTTGTTCGA	900
AGGTACTTAT	TCTATTTAAA	ACTTGATATA	ATTCTTGAGT	TATATCTAAT	TTATCTAATT	960
CACTTGTCCC	TGGAGTAGTT	TGATCTAATT	GATGTAATT	ATTTAATAAA	TCACGTTCTA	1020
AATCAGTTG	TTTAGATAAA	TCATTTAATA	AATCATCTTC	AGCATT		1066

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 302 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

45 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

50 Met	Pro	Arg	Phe	Lys	Gln	Val	Asp	Val	Phe	Thr	Asn	Val	Lys	Tyr	Leu
1															
														10	15
Gly Asn Pro Val Ala Val Ile Tyr Asp Ser Asp Asn Leu Thr Thr Gln															

EP 0 982 401 A2

	20	25	30
	Glu Met Gln Lys Ile Ala Arg Trp Thr Asn Leu Ser	lu Thr Thr Phe	
5	35 40	45	
	Ile Leu Thr Pro Lys Ser Ser Ile Ala Xaa Tyr Ser Ile Arg Ile Phe		
	50 55	60	
	Thr Ser Gly Gly Asn Glu Leu Pro Phe Ala Gly His Pro Thr Leu Gly		
	65 70	75	80
10	Thr Ala Phe Ala Leu Leu Glu Asp Gly Lys Ile Lys Pro Asn Asp Asn		
	85 90	95	
	Gly Gln Ile Ile Gln Glu Cys Gly Ala Gly Leu Val Lys Ile Ser Val		
	100 105	110	
	Glu Lys Thr Pro Asn Asn Asn Ser Asn Glu Leu Pro Phe Leu Leu Ser		
	115 120	125	
15	Phe Glu Leu Pro Tyr Phe Lys Phe His Glu Ile Asp Asp Lys Val Ile		
	130 135	140	
	Glu Glu Leu Gln His Ser Trp Asn Gly Thr Asn Ile Ile Gly Lys Pro		
	145 150	155	160
20	Val Leu Ile Asp Ala Gly Pro Lys Trp Ala Val Phe Gln Leu Gly Ser		
	165 170	175	
	Gly Lys Glu Val Leu Asp Leu Asn Xaa Asp Leu Ala Gln Ile Glu Arg		
	180 185	190	
25	Leu Ser Leu Glu Asn Gly Trp Thr Gly Ile Gly Val Phe Gly Lys His		
	195 200	205	
	Asn Glu Asn Gly Asp Ser Val Glu Leu Arg Asn Ile Ala Pro Ala Val		
	210 215	220	
30	Gly Val Ala Glu Asp Pro Ala Cys Gly Ser Gly Ser Gly Ala Ile Gly		
	225 230	235	240
	Ala Tyr Leu Ala Asn His Val Phe Asn Glu Lys Glu Lys Phe Thr Ile		
	245 250	255	
	Asp Ile Ser Gln Gly Lys Pro Ile Glu Arg Asp Ala Lys Ile Gln Val		
	260 265	270	
35	Lys Val Asn Arg Leu Thr Thr Lys Asn Gly Asp Leu Ser Ile His Val		
	275 280	285	
	Gly Gly His Ala Ile Thr Cys Phe Glu Gly Thr Tyr Ser Ile		
	290 295	300	

40 (2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2829 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATGACGGAAA CTGTGATAGA AAAGAAAAGA AAGGTTGATT TAAATGCCTC AGGTATTACA

60

EP 0 982 401 A2

	AAACAACCAA AAGCTTCTAA AATCTTCAGT CCATTCAGAG TTTAGGAA TGTTACAGAC	120
	TCAACTCCTT TTGCCATGGG GACATTAGGT TCAACATTTT ATGCTGTAC TTCTGTTGGC	180
5	AGATCTTCC AAATTTATGA CTTGGCTACA TTACATTAT TGTGTTTC CCAAACCTAA	240
	ACTCCTCAA GAATTACAAG TTGGCTGCA CACCATCACT ATGCTATGC ATCTTATGGT	300
	GATCGTATTG GTATTTTAG ACGTGGTAGA TTAGAGCATG AATTGGTTG TGAAGGAAAC	360
10	TCTACAGTTA ACCAATTATT AGTATTTGGA GAATACCTTA TTGCTACAC ATTAGAAGGT	420
	GATATTTCG TATTTAGAAA AACTGAAGGA AAGAAATTCC CAACTGAATT ATACACTACA	480
	ATCAGAATAA TTAATTCTT AGTTGAAGGA GAATTGTGG GATTAATTCA TCCACCTACG	540
	TATTTAAATA AAGTAATTGT TGCTACTACT CAATCTGTGT TTGTTATAAA TGTGAGAACT	600
15	GGCAAATTAT TATACAAATC CCGGGAATTAA CAATCGAAG GCGAARRAGAT TTCACTCAATC	660
	GAAGCTGCTC CAGTTTGGA TGTAATTGCT GTGGTACAT CTAATGGAAA TGTATTTTA	720
	TTCAACATTA AAAAGGGAA AGTGTGGGC CAAAAAATTA TTACTTCTGG AACTGAATCT	780
	TCTTCGAAAG TTGCCTCGAT CTCTTTAGA ACAGATGGAG CACCTCATTG GTTGCTGGT	840
20	TTGAATAACG GGGACTTATA TTCTACGAT TTAGACAAAG AATCACGTGT TCATGTTTG	900
	AGAAATGCCC ATAAAGAGAC TCATGGGGGT GTGCAAACG CCAAATTCTT GAATGGTCAA	960
	CCAATAGTAT TATCAAATGG TGGTGATAAT CATTGAAAG AATTGTTTT TGATCCTAAT	1020
25	TTAACCACTT CGAATTCACTC CATTGTTCCCT CCTCCAAGAC ATCTCAGATC TAGAGGTGGG	1080
	CATTCACTAC CACCACTAGC TATTGAATT CCTCAAGAAG ATAAAACCCA TTTTTATTG	1140
	AGTGCCTCTA GAGATAAAC ATTTGGACA TTCTCTTGA GAAAAGATGC TCAAGCACAG	1200
	GAAATGTCTC AAAGATTGCA AAAATCTAAG GATGGTAAAA GACAGGCTGG ACAAGTTGTT	1260
30	TCTATGAGAG AGAAATTCCC AGAAATCATT TCCATTTCAT CCTCTTATGC CAGAGAAGGT	1320
	GATTGGAAA ATATCATAAC CGCCCAACAAG GATGAAACTT TTGCGAGAAC ATGGGATTCA	1380
	AGAAATAAAA GACTCGGTAG ACATTTGTTA AACACTATTG ATGGTGGCAT TGTGAAATCT	1440
35	GTATGTGTGT CTCAGTGTGG TAATTTGGT TTAGTGGAT CATCACTGGG TGGTATTGGA	1500
	TCATACAAACC TTCAAAGTGG ATTGTTGCGT AAAAATATG TTTACATAA ACAAGCTGTC	1560
	ACCGGTTTAG CAATTGATGG AATGAATAGA AAAATGGTTA GTTGTGGTT AGATGGAATT	1620
	GTGGGATTCT ATGATTTGG AAAGTCTGTC TATTTAGGCA AATTACAAC TGAAGCACCT	1680
40	ATAACATCCA TGATATATCA CAAACTGTCT GATCTTGTG CTTGTGCCTT GGATGATTG	1740
	TCCATAGTTG TTATTGACGT GACTACTCAA AAAGTCATAA GAATATTATA TGGTCATACC	1800
	AACAGAAATT CAGGAATGGA TTCTCGCCT GATGGGAGAT GGATAGTTTC AGTTGCATTG	1860
45	GACTCCACTT TGCAGACTTG GGACTTGCCTA ACTGGTGGT GTATTGATGG GGTGATTTA	1920
	CCAATTGTGG CAACTGCAGT TAAATTTCT CCTATTGGTG ATATCTTAGC GACAACACAT	1980
	GTCTCTGGAA ATGGTGTATC CTTATGGACT AATCGTGCCT AGTCAGGCC TGTGTCCACC	2040
	AGACACGTAG AAGAAGATGA GTTTCAACT ATTTTATTAC CAAATGCTTC TGGAGATGGC	2100
50	GGTTCAACAA TGCTAGACGG GTTTTGGAC GAGGATTCTA ATGAAGACGG CACTATTGAT	2160
	GAACAGTATA CATCTGCTGC TCAAATTGAT GCATCCTGAA TTACTTTATC ATCAGAGCCA	2220

EP 0 982 401 A2

5	AGATCAAAAT TCAACACTT ATTGCATTT GATACCATTA AACARCAAAG CAAACCGAAA	2280
	GAAGCACCTA AAAACCCAGA AAATGCACCT TTCTTTTAC AATTGACTGG ACAAGCAGTT	2340
	GGTGATAGGG CATCGGTTGC TGAAGGCAAAC TTTCAGAAC AAACAAATAA CACTGTTGAA	2400
	GAAACCAACA GCAAATTGGT ACAAAACGCTA ACCACGCATT TGAAAGTGAA	2460
10	TTCACAAAAC TATTAAGGGA AGCTGGAGAG AGTGGACAAAT TTGAAAGATT TTTGACTTAC	2520
	TTACTTAAC TATCTCCTGC TGTATTGGAC TTGGAAATTA GATCACTTAA TTCATTGTT	2580
	CCATTGACTG AAATGACAAA TTTTATTCAA GCTTTAAATG CTGGTTGAA ATCAAACGCA	2640
	AATTATGAAA TATGGGAAAC TTTATATGCC ATGTTTTCA ACATACATGG TGATGTTATC	2700
	CATCAGTTG AAAATGAAAC TAGTCTTCAT GAAGCTTTGG AAGAATACAG ACAGTTAAAT	2760
15	GATGAAAAGA ATAACAAAAT GGATTCTTTA GTGAAATATT GTGCTAGTAT CGTAAGTTT	2820
	ATTAGTTAG	2829

(2) INFORMATION FOR SEQ ID NO: 14:

20	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 942 amino acids
	(B) TYPE: amino acid
	(C) STRANDEDNESS:
	(D) TOPOLOGY: unknown
	(ii) MOLECULE TYPE: peptide
25	(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

30	Met Thr Glu Thr Val Ile Glu Lys Lys Arg Lys Val Asp Leu Asn Ala	1 5 10 15
	Ser Gly Ile Thr Lys Gln Pro Lys Ala Ser Lys Ile Phe Ser Pro Phe	20 25 30
	Arg Val Leu Gly Asn Val Thr Asp Ser Thr Pro Phe Ala Met Gly Thr	35 40 45
35	Leu Gly Ser Thr Phe Tyr Ala Val Thr Ser Val Gly Arg Ser Phe Gln	50 55 60
	Ile Tyr Asp Leu Ala Thr Leu His Leu Leu Phe Val Ser Gln Thr Gln	65 70 75 80
40	Thr Pro Ser Arg Ile Thr Ser Leu Ala Ala His His His Tyr Val Tyr	85 90 95
	Ala Ser Tyr Gly Asp Arg Ile Gly Ile Phe Arg Arg Gly Arg Leu Glu	100 105 110
45	His Glu Leu Val Cys Glu Gly Asn Ser Thr Val Asn Gln Leu Leu Val	115 120 125
	Phe Gly Glu Tyr Leu Ile Ala Thr Thr Leu Glu Gly Asp Ile Phe Val	130 135 140
50	Phe Arg Lys Thr Glu Gly Lys Lys Phe Pro Thr Glu Leu Tyr Thr Thr	145 150 155 160
	Ile Arg Ile Ile Asn Ser Leu Val Glu Gly Glu Ile Val Gly Leu Ile	165 170 175

EP 0 982 401 A2

	His Pro Pro Thr Tyr Leu Asn Lys Val Ile Val Ala Thr Thr Gln Ser			
	180	185	190	
	Val Phe Val Ile Asn Val Arg Thr Gly Lys Leu Leu Tyr Lys Ser Arg			
5	195	200	205	
	Glu Leu Gln Phe Glu Gly Glu Lys Ile Ser Ser Ile Glu Ala Ala Pro			
	210	215	220	
	Val Leu Asp Val Ile Ala Val Gly Thr Ser Asn Gly Asn Val Phe Leu			
10	225	230	235	240
	Phe Asn Ile Lys Lys Gly Lys Val Leu Gly Gln Lys Ile Ile Thr Ser			
	245	250	255	
	Gly Thr Glu Ser Ser Ser Lys Val Ala Ser Ile Ser Phe Arg Thr Asp			
	260	265	270	
15	Gly Ala Pro His Leu Val Ala Gly Leu Asn Asn Gly Asp Leu Tyr Phe			
	275	280	285	
	Tyr Asp Leu Asp Lys Lys Ser Arg Val His Val Leu Arg Asn Ala His			
	290	295	300	
20	Lys Glu Thr His Gly Gly Val Ala Asn Ala Lys Phe Leu Asn Gly Gln			
	305	310	315	320
	Pro Ile Val Leu Ser Asn Gly Gly Asp Asn His Leu Lys Glu Phe Val			
	325	330	335	
	Phe Asp Pro Asn Leu Thr Thr Ser Asn Ser Ser Ile Val Pro Pro Pro			
25	340	345	350	
	Arg His Leu Arg Ser Arg Gly Gly His Ser Ala Pro Pro Val Ala Ile			
	355	360	365	
	Glu Phe Pro Gln Glu Asp Lys Thr His Phe Leu Leu Ser Ala Ser Arg			
	370	375	380	
30	Asp Lys Thr Phe Trp Thr Phe Ser Leu Arg Lys Asp Ala Gln Ala Gln			
	385	390	395	400
	Glu Met Ser Gln Arg Leu Gln Lys Ser Lys Asp Gly Lys Arg Gln Ala			
	405	410	415	
	Gly Gln Val Val Ser Met Arg Glu Lys Phe Pro Glu Ile Ile Ser Ile			
35	420	425	430	
	Ser Ser Ser Tyr Ala Arg Glu Gly Asp Trp Glu Asn Ile Ile Thr Ala			
	435	440	445	
	His Lys Asp Glu Thr Phe Ala Arg Thr Trp Asp Ser Arg Asn Lys Arg			
	450	455	460	
40	Val Gly Arg His Leu Leu Asn Thr Ile Asp Gly Gly Ile Val Lys Ser			
	465	470	475	480
	Val Cys Val Ser Gln Cys Gly Asn Phe Gly Leu Val Gly Ser Ser Ser			
	485	490	495	
45	Gly Gly Ile Gly Ser Tyr Asn Leu Gln Ser Gly Leu Leu Arg Lys Lys			
	500	505	510	
	Tyr Val Leu His Lys Gln Ala Val Thr Gly Leu Ala Ile Asp Gly Met			
	515	520	525	
	Asn Arg Lys Met Val Ser Cys Gly Leu Asp Gly Ile Val Gly Phe Tyr			
50	530	535	540	
	Asp Phe Gly Lys Ser Val Tyr Leu Gly Lys Leu Gln Leu Glu Ala Pro			
	545	550	555	560

EP 0 982 401 A2

Ile Thr Ser Met Ile Tyr His Lys Ser Ser Asp Leu Val Ala Cys Ala
 565 570 575
 Leu Asp Asp Leu Ser Ile Val Val Ile Asp Val Thr Thr Gln Lys Val
 580 585 590
 5 Ile Arg Ile Leu Tyr Gly His Thr Asn Arg Ile Ser Gly Met Asp Phe
 595 600 605
 Ser Pro Asp Gly Arg Trp Ile Val Ser Val Ala Leu Asp Ser Thr Leu
 610 615 620
 10 Arg Thr Trp Asp Leu Pro Thr Gly Gly Cys Ile Asp Gly Val Ile Leu
 625 630 635 640
 Pro Ile Val Ala Thr Ala Val Lys Phe Ser Pro Ile Gly Asp Ile Leu
 645 650 655
 Ala Thr Thr His Val Ser Gly Asn Gly Val Ser Leu Trp Thr Asn Arg
 15 660 665 670
 Ala Gln Phe Lys Pro Val Ser Thr Arg His Val Glu Glu Asp Glu Phe
 675 680 685
 Ser Thr Ile Leu Leu Pro Asn Ala Ser Gly Asp Gly Ser Thr Met
 690 695 700
 20 Leu Asp Gly Phe Leu Asp Glu Asp Ser Asn Glu Asp Gly Thr Ile Asp
 705 710 715 720
 Glu Gln Tyr Thr Ser Ala Ala Gln Ile Asp Ala Ser Leu Ile Thr Leu
 725 730 735
 Ser Ser Glu Pro Arg Ser Lys Phe Asn Thr Leu Leu His Leu Asp Thr
 25 740 745 750
 Ile Lys Gln Gln Ser Lys Pro Lys Glu Ala Pro Lys Lys Pro Glu Asn
 755 760 765
 Ala Pro Phe Phe Leu Gln Leu Thr Gly Gln Ala Val Gly Asp Arg Ala
 30 770 775 780
 Ser Val Ala Glu Gly Lys Thr Ser Glu Gln Thr Asn Asn Thr Val Glu
 785 790 795 800
 Glu Thr Asn Ser Lys Leu Arg Lys Leu Asp Thr Asn Gly Asn His Ala
 805 810 815
 35 Phe Glu Ser Glu Phe Thr Lys Leu Leu Arg Glu Ala Gly Glu Ser Gly
 820 825 830
 Gln Phe Glu Arg Phe Leu Thr Tyr Leu Leu Asn Leu Ser Pro Ala Val
 835 840 845
 40 Leu Asp Leu Glu Ile Arg Ser Leu Asn Ser Phe Val Pro Leu Thr Glu
 850 855 860
 Met Thr Asn Phe Ile Gln Ala Leu Asn Ala Gly Leu Lys Ser Asn Ala
 865 870 875 880
 45 Asn Tyr Glu Ile Trp Glu Thr Leu Tyr Ala Met Phe Phe Asn Ile His
 885 890 895
 Gly Asp Val Ile His Gln Phe Glu Asn Glu Thr Ser Leu His Glu Ala
 900 905 910
 Leu Glu Glu Tyr Arg Gln Leu Asn Asp Glu Lys Asn Asn Lys Met Asp
 50 915 920 925
 Ser Leu Val Lys Tyr Cys Ala Ser Ile Val Ser Phe Ile Ser
 930 935 940

EP 0 982 401 A2

(2) INFORMATION FOR SEQ ID NO: 15:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 725 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

AACCTGGCAA	TTAACTGCC	GGCAAGTGAT	ACCAAGGAGAT	AGGTGTGAT	AGATTATAAT	60
GGAACGCCGA	TTTTGCAGT	ATCACCGCTA	ATAAGGACAG	CAGTTGGACA	TCGGTACATG	120
AGAGAGCAAT	GTAAGTCTTG	ATAGTAATGA	GCCGTGTTGA	AGTAGTATT	TAATCTAATT	180
TTACTCAAAA	AAGGACAATG	GAGATCTGGA	GATAACAGCA	CACTAATCGG	TTCTAGACAT	240
AGACTAAGCC	TGAAAGGGGG	TACTACAGCT	TGTTTGAAA	AGGTTGCGT	TGTATAGGCA	300
GTTAAATGTG	TGTTTTTTT	GGGTAGAATT	TGAGAAAAAG	TTGACTGAAA	AAAATGCAAG	360
AAACGGGGTG	ATCATGAAA	TAGACACACA	CAAAAGTCA	AAAAACAATG	GAAAAGCTTC	420
AGAATAAGCA	GTAGGAGGTG	TCTGAATTGA	GTTGTATTG	TTATTTAGAG	TTTTAAATTA	480
GAGTTGAAA	TTTTGGGT	GAATTTACGA	AAAAGTCGAA	CAAAAAAACG	ACAAGTCAGG	540
GTGATTGCAA	AAAACAGAA	ACAATAGATA	ATCTTAAATT	AAGGTAGTAG	AGGCTCTGTG	600
AAGTAATTAA	GACTTTAAC	ACGGGGGCAC	GAGTCAGTGT	TAGAGTTGTG	AAGTTTATT	660
GGCTAGTGAA	TTGACTGGCA	AGATTGTTAA	ACGTGGGT	GAAGGACA	ACGCATCGAC	720
AGGTT						725

(2) INFORMATION FOR SEQ ID NO: 16:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1144 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

40 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CCATGATATA	GAAATTGGTG	GGTCAACGTA	CTATCAAATT	AACATAAAAC	TACCACTTCG	60
GTCATTCACT	ATAAAGAAAC	GGTACCTGGA	ATTCCAGCAA	TTGGTGCTGG	ACTTGAGTCG	120
TAATCTAGGC	ATTGATAGTC	GAGATTTCC	ATATGAATT	CCTGGGAAAC	GGATCAACTG	180
GCTTAACAAG	ACCAAGTATTG	TTGAGGAGAG	AAAAGTGGGA	CTTGCAGAAT	TTCTCAATAA	240
CCTCATTCAA	GACTCAACAC	TTCAGAATGA	ACGAGAAGTG	TTGTCGTTT	TGCAATTGCC	300
GTCTAATTAA	AGATTCCACCA	AGGATATGTT	ACAGAATAAT	CGAGCAGACT	TGGATTCTGT	360

EP 0 982 401 A2

5	GCAAAATAAC TGGTACGATG TATATCGTAA GTTGAAACTG GATATACTCA ACGAATCGTC	420
	TAGCAGCATT AGTGAACAGA TACATATTG TGATCGCATT AGTCGGGTCT ACCAACCCAG	480
	GATTCTCGAC TTGGTCAGGG CTATTGGTAC AGATAAAAGAA GAGGCCCTAA AGAAGAAGCA	540
	GTTGGTTTC CAATTACAAG AGAGTATAGA TAATTGTTA GTACAGGAAG TTCCCCGATC	600
	AAAGAGGGTG TTGGGTGGAG CAGTTAAGGA AACGCCAGAG ACATTACCAT TAAACAATAA	660
10	AGAACTTCTT CAACACCAAG TACAAATTCA TCAAAACCAA GACAAAGAAC TAGACCAGCT	720
	TAGGGTGTAA ATTGCCCGGC AGAAACAGAT TGGCGAGCTA ATTAATGCAAG AAGTAGAGGA	780
	ACAGRATGAA ATGTTGGATA GGTTAATGA AGAGGTCGAC TACACGTCCA GCAAAATCAA	840
	GCAAGCAAGA CGCAGAGCTA AGAAGATATT ATAGTAATTG TTTCGCTACT TCGATATTAT	900
15	CTGCCATTGA CGTTATTCTT GCAGGTTGGC CCAATTGTTTC GTTGAAAGT TTTCGAGGT	960
	CTTCAGCGTC TAATGCCCTA TCTGAGCTCT CGCCATCGAG TTTCCAAAC CCGCCGATAT	1020
	TTTGAAAGAA TCTTGAAATG CCAAACCGTC GTGGCGGGAA CGATCTGCCCT GCGTTGGCCA	1080
	AGTTGAATAT GCTAGGGTGG TACTGTAAAT AGAAGACAGA TCCAATAAAC GTTCCTATAA	1140
20	ATGC	1144

(2) INFORMATION FOR SEQ ID NO: 17:

25	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 290 amino acids	
	(B) TYPE: amino acid	
	(C) STRANDEDNESS:	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: peptide	
30	(iii) HYPOTHETICAL: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:	
35	His Asp Ile Glu Ile Gly Gly Ser Thr Tyr Tyr Gln Ile Asn Ile Lys	
	1 5 10 15	
	Leu Pro Leu Arg Ser Phe Thr Ile Lys Lys Arg Tyr Ser Glu Phe Gln	
	20 25 30	
	Gln Leu Val Ser Asp Leu Ser Arg Asn Leu Gly Ile Asp Ser Arg Asp	
	35 40 45	
40	Phe Pro Tyr Glu Leu Pro Gly Lys Arg Ile Asn Trp Leu Asn Lys Thr	
	50 55 60	
	Ser Ile Val Glu Glu Arg Lys Val Gly Leu Ala Glu Phe Leu Asn Asn	
	65 70 75 80	
	Leu Ile Gln Asp Ser Thr Leu Gln Asn Glu Arg Glu Val Leu Ser Phe	
45	85 90 95	
	Leu Gln Leu Pro Ser Asn Phe Arg Phe Thr Lys Asp Met Leu Gln Asn	
	100 105 110	
	Asn Arg Ala Asp Leu Asp Ser Val Gln Asn Asn Trp Tyr Asp Val Tyr	
50	115 120 125	
	Arg Lys Leu Lys Ser Asp Ile Leu Asn Glu Ser Ser Ser Ile Ser	
	130 135 140	

EP 0 982 401 A2

Glu Gln Ile His Ile Arg Asp Arg Ile Ser Arg Val Tyr Gln Pro Arg
 145 150 155 160
 Ile Leu Asp Leu Val Arg Ala Ile Gly Thr Asp Lys Glu Glu Ala Leu
 165 170 175
 Lys Lys Lys Gln Leu Val Ser Gln Leu Gln Glu Ser Ile Asp Asn Leu
 180 185 190
 Leu Val Gln Glu Val Pro Arg Ser Lys Arg Val Leu Gly Gly Ala Val
 195 200 205
 Lys Glu Thr Pro Glu Thr Leu Pro Leu Asn Asn Lys Glu Leu Leu Gln
 210 215 220
 His Gln Val Gln Ile His Gln Asn Gln Asp Lys Glu Leu Asp Gln Leu
 225 230 235 240
 Arg Val Leu Ile Ala Arg Gln Lys Gln Ile Gly Glu Leu Ile Asn Ala
 245 250 255
 Glu Val Glu Glu Gln Asn Glu Met Leu Asp Arg Phe Asn Glu Glu Val
 260 265 270
 Asp Tyr Thr Ser Ser Lys Ile Lys Gln Ala Arg Arg Arg Ala Lys Lys
 275 280 285
 Ile Leu
 290

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2736 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:11
(D) OTHER INFORMATION:/note= "N = G or A or T or C"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:2723..2724
(D) OTHER INFORMATION:/note= "N = A or T or C or G"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:2714..2715
(D) OTHER INFORMATION:/note= "N = A or T or C or G"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:2710
(D) OTHER INFORMATION:/note= "N = A or T or C or G"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:2706..2707
(D) OTHER INFORMATION:/note= "N = A or T or C or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

ATGGAAAAAA NTTGGCGAG TGTAAAGTTG TACACCGATT TGGAGTGTGT TTTTAATTCA

60

EP 0 982 401 A2

	AACTATCCAA CAAGAATTGT TTGGGGTGC TCTTACAATT TTGGAATTCA ACAGATGATG	120
	GCAAACTTG ATCGGTTTC AAAACCACCA GTGGATCCAT CTACAAAATT AGGATTTGG	180
5	GATAAGTTAA ACTATATCTT ACATGGTAA TGCCAAATCA GAACTAGGAA AAGTTAGAA	240
	GTTGCATTTA ARGGATCAAG AGATCCGTAT GATTTGTTCA CGACTGCAGG CGGGTTGTA	300
	TTGTCATTTA GAAAGAATGT TGTCTGGAC ATCAATAAG ACGATAATTC GAAAATTAC	360
10	TTCGATATCA CGGCAGATAA AGTTCCCTGG TATATTCCAA ACTATTTAGC AGGACCATTA	420
	TTGGCTTGGA CAAGAAGTAG TAAAAATTCA ATTTATTTAC CAAATTCAAC AATGTGGTT	480
	AATTCTGCT TTGCATATTA CCTCAAGAT TTTACTGGAC AAGCTGATTT TGATCATGCT	540
	GCCCCAGTAT TTGAAAGAAA TGTGGTCAAT CTTAGTGGAG GAATTCAATT TCAAGTTGGG	600
15	TTTCTACTTG AACGTAAGA TACAAATGGT AAGAGAACCG ATGAATTCAA ACCTCATTAC	660
	GAAGTGCAGT TGTGATCC CAAGTATTGT GAGAAGGAC ATGACTCTTA TGCTGGTT	720
	CGAAGTCAAT TTATACATAT GGCTATCTCA TTGGARTCAA CAAACAGTTC AAGTTATAAT	780
	ACAATCCATC TTAGTCCTGG TACTTTCCAA CAGTTTTCG ATTGGTGAA GTTATTGCT	840
20	AGTAATATGC AGTTACCTAT TAGACGTGGC AAAATGTTT GAGAAGCAGA AGAATCTGTC	900
	AAGTTTTCGC AACATTTATT CACAAAACAAG TTTCTTTCA TGTTGAATC TTTGTTTATT	960
	GCTCATGTT ATCGAGACGA AATTGTTGAT ATCAATAACG ATAGAATAGA AAGTATTGGT	1020
	TTAAGAGCCA AAGTAGATGA TTTTATGGTT GATTTACATC AAAGAAAAGA GCCAGCAACC	1080
25	CTTTACCATG AAGAATTATC TAAGAATGAG AAGGTGATGA AAATGAATT TGATTTAGGA	1140
	GAAGTCGTT TATCAGGAAT AGACTTACGT GTCATGCATG TTTCATTCT CCAAAATTAA	1200
	TACACTCAAT CACATTCCAA TTCAGGTGAC GCTAAATCAA CTTATAATT TTACGACAAT	1260
30	GATCATCGAT GGTTGATAT TATGGATTTC CAAGAGGCAT TTTGACATC AATTAAGGAT	1320
	TGTGTCAGGA CAGTTGATAT TTATCCATTG ATGTATTAC AAAGATTCTT TTATGAAAGA	1380
	GATACACATG GTGGCAACTC TGAGGATGAG ACTGCATTG GAAAAGAAGT TATTCAAAA	1440
35	TGTAATTGG GTGCCATGAA TCCCTGGAA ACAAGATTGA ATGTATTGGT TCAAAGACTT	1500
	AACGCTCTAC AAGAACAAAGT CAAAAAATTG TCCAAAACAT CTGCTCCAGA ACCTGTAGCA	1560
	GATTTGAAAA AACGAATTCT GTTTTGCAA AAAGAGATTA CCACAAACCAA AGCTGGCGTT	1620
	AAGTCGAAAA TGCCTCGTAC ATCCACTATA AATGGTATGA ATAATTCTGA AAATTACAC	1680
40	AATAAGTTA CTTCTATAA CATGCTTCTT AAATGGAATT TCAATTGTCG GAATTGACA	1740
	TTGAAATACA TACATTTGT GAAATTGAAA TCACAACCTC GAAATTACTT GTCACACAAAG	1800
	TCCATTGAAA CACTTGAAA AATGATGGAT AGTGTAAATG CATAACACGA TAAGGACGAT	1860
	TTGTCATCGA CGTCAGAATT AATCCGTGT TTCACACTGG AAGGGGTTAA ATCACAGACA	1920
45	TCTACCAAGCA ARGATATCAC TTCACAACAG AAACCTGACA ATTTCAACAC AATATTACGA	1980
	GAGACCGAC CAGACGAAAA AGTGGTTGAG GATTATTGTA TTGACGTGAT CGCACCTCAA	2040
	ATTCAATTAC AAAGTGAGGA TTATCCTGAT TCTGTTGTGC TCATCTCTAC ACCATCTATT	2100
50	AAAGGTAAAA TTTGTCCAT TAGGGATTCC AGGAATAATG CAAACCAAAT CTTGTTAGAA	2160
	ACTAGGTATG GTATTTTACT AAAAGATGCC AATGTTTTG TATTAACAA AGAGGATATT	2220

EP 0 982 401 A2

5	GTAGGGTGTG CAGATAT TT AAGTATTAGT AATCCATATG GAGCTAAATC TAATTGGCCA	2280
	CCATGGCTAG GAACAGAAAT AACCCAAAAT GGTAATGGG CTGGAGCCAA CAACTTATTG	2340
	ATTGAAAAGC TTCTGTTAT GACAATGTGT TATGAAAGTG AAATTTGTC AAGCAAGCTT	2400
	TCTCCAAATG CACAAGATCT GGATCAAGAA GAGCAAGAAA ATTACAAATGA TGATAATTG	2460
	AAACAGGCTC CTCTTCGACT TGGTATTGAT ATGCCCTCTG TGGTGATTAC ATCTACATCA	2520
10	AGTCAAACTA TTACCTTATA TGTTATCATA GTGAGCTTGT TGTTTATAG CGAGCCTATG	2580
	AGTAAAGTGA TCCACAAGAA AATCGAAAAG ATGAAGTTT CTATTGATTT CGAAGATTTG	2640
	GGTGCTCTTA CTAGCAGATT AACGAAAATG CAGCAACATC ATAAAATGTT GAAAGTATTG	2700
	TCTAANNACN AATNNNTTCC CGNNCGGGGG AATTAA	2736

15 (2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 911 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

	Met Glu Lys Xaa Leu Ala Ser Val Lys Leu Tyr Thr Asp Leu Glu Cys	
	1 5 10 15	
30	Val Phe Asn Ser Asn Tyr Pro Thr Arg Ile Val Trp Gly Ala Ser Tyr	
	20 25 30	
	Asn Phe Gly Ile Gln Gln Met Met Ala Asn Phe Asp Arg Phe Ser Lys	
	35 40 45	
35	Pro Pro Val Asp Pro Ser Thr Lys Leu Gly Phe Trp Asp Lys Leu Lys	
	50 55 60	
	Tyr Ile Leu His Gly Lys Cys Gln Ile Arg Thr Arg Lys Ser Leu Glu	
	65 70 75 80	
	Val Ala Phe Lys Gly Ser Arg Asp Pro Tyr Asp Leu Phe Thr Thr Ala	
	85 90 95	
40	Gly Gly Phe Val Leu Ser Phe Arg Lys Asn Val Val Trp Asp Ile Asn	
	100 105 110	
	Lys Asp Asp Asn Ser Lys Asn Tyr Phe Asp Ile Thr Ala Asp Lys Val	
	115 120 125	
45	Ser Trp Tyr Ile Pro Asn Tyr Leu Ala Gly Pro Leu Leu Ala Trp Thr	
	130 135 140	
	Arg Ser Ser Lys Asn Ser Ile Tyr Leu Pro Asn Ser Pro Asn Val Val	
	145 150 155 160	
50	Asn Ser Cys Phe Ala Tyr Tyr Leu Gln Asp Phe Thr Gly Gln Ala Asp	
	165 170 175	
	Phe Asp His Ala Ala Arg Val Phe Glu Arg Asn Val Val Asn Leu Ser	
	180 185 190	

EP 0 982 401 A2

Gly Gly Ile His Phe Gln Val Gly Phe Leu Leu Glu Arg Lys Asp Thr
 195 200 205
 Asn Gly Lys Arg Thr Asp Glu Phe Lys Pro His Tyr Glu Val Gln Leu
 210 215 220
 5 Phe Asp Pro Lys Tyr Cys Glu Lys Gly His Asp Ser Tyr Ala Gly Phe
 225 230 235 240
 Arg Ser Gln Phe Ile His Met Ala Ile Ser Leu Glu Ser Thr Asn Ser
 245 250 255
 10 Ser Ser Tyr Asn Thr Ile His Leu Ser Pro Gly Thr Phe Gln Gln Phe
 260 265 270
 Phe Asp Trp Trp Lys Leu Phe Ala Ser Asn Met Gln Leu Pro Ile Arg
 275 280 285
 15 Arg Gly Lys Met Phe Gly Glu Ala Lys Glu Ser Val Lys Phe Ser Gln
 290 295 300
 His Leu Phe Thr Asn Lys Phe Ser Phe Met Leu Lys Ser Leu Phe Ile
 305 310 315 320
 Ala His Val Tyr Arg Asp Glu Ile Val Asp Ile Asn Asn Asp Arg Ile
 20 325 330 335
 Glu Ser Ile Gly Leu Arg Ala Lys Val Asp Asp Phe Met Val Asp Leu
 340 345 350
 His Gln Arg Lys Glu Pro Ala Thr Leu Tyr His Glu Glu Leu Ser Lys
 355 360 365
 25 Asn Glu Lys Val Met Lys Met Asn Phe Asp Leu Gly Glu Val Val Leu
 370 375 380
 Ser Gly Ile Asp Leu Arg Val Met His Val Ser Phe Leu Gln Asn Leu
 385 390 395 400
 30 Tyr Thr Gln Ser His Ser Asn Ser Gly Asp Ala Lys Ser Thr Tyr Asn
 405 410 415
 Ile Tyr Asp Asn Asp His Arg Trp Phe Asp Ile Met Asp Phe Gln Glu
 420 425 430
 Ala Phe Leu Thr Ser Ile Lys Asp Cys Val Arg Thr Val Asp Ile Tyr
 435 440 445
 35 Pro Leu Met Tyr Leu Gln Arg Phe Phe Tyr Glu Arg Asp Thr His Gly
 450 455 460
 Gly Lys Ser Glu Asp Glu Thr Ala Phe Gly Lys Glu Val Ile His Lys
 465 470 475 480
 40 Cys Asn Leu Gly Ala Met Asn Pro Leu Glu Thr Arg Leu Asn Val Leu
 485 490 495
 Val Gln Arg Leu Asn Ala Leu Gln Glu Gln Val Lys Lys Leu Ser Lys
 500 505 510
 45 Thr Ser Ala Pro Glu Pro Val Ala Asp Leu Lys Lys Arg Ile Ser Phe
 515 520 525
 Leu Gln Lys Glu Ile Ser Thr Thr Lys Ala Gly Val Lys Ser Lys Met
 530 535 540
 Arg Arg Thr Ser Thr Ile Asn Gly Met Asn Asn Ser Glu Asn Tyr His
 545 550 555 560
 50 Asn Lys Phe Thr Phe Tyr Asn Met Leu Leu Lys Trp Asn Phe Asn Cys
 565 570 575

EP 0 982 401 A2

Arg Asn Leu Thr Leu Lys Tyr Ile His Phe Val Lys Leu Lys Ser Gln
 580 585 590
 Leu Arg Asn Tyr Leu Ser His Lys Ser Ile Glu Thr Leu Glu Lys Met
 595 600 605
 5 Met Asp Ser Val Asn Ala Tyr Asn Asp Lys Asp Asp Leu Ser Ser Thr
 610 615 620
 Ser Glu Ile Ile Arg Arg Phe Thr Ser Glu Gly Val Lys Ser Gln Thr
 625 630 635 640
 10 Ser Thr Ser Lys Asp Ile Thr Ser Gln Gln Lys Leu Asp Asn Phe Asn
 645 650 655
 Thr Ile Leu Arg Glu Thr Arg Pro Asp Glu Lys Val Val Glu Asp Tyr
 660 665 670
 Leu Ile Asp Val Ile Ala Pro Gln Ile Gln Leu Gln Ser Glu Asp Tyr
 15 675 680 685
 Pro Asp Ser Val Val Leu Ile Ser Thr Pro Ser Ile Lys Gly Lys Ile
 690 695 700
 Leu Ser Ile Arg Asp Ser Arg Asn Asn Ala Asn Gln Ile Leu Leu Glu
 705 710 715 720
 20 Thr Arg Tyr Gly Ile Leu Leu Lys Asp Ala Asn Val Phe Val Leu Asn
 725 730 735
 Lys Glu Asp Ile Val Gly Cys Pro Asp Met Leu Ser Ile Ser Asn Pro
 740 745 750
 25 Tyr Gly Ala Lys Ser Asn Trp Pro Pro Trp Leu Gly Thr Glu Ile Thr
 755 760 765
 Gln Asn Gly Lys Trp Ala Gly Ala Asn Asn Leu Leu Ile Glu Lys Leu
 770 775 780
 Ser Val Met Thr Met Cys Tyr Glu Ser Glu Ile Leu Ser Ser Lys Leu
 30 785 790 795 800
 Ser Pro Asn Ala Gln Asp Ser Asp Gln Glu Glu Gln Glu Asn Tyr Asn
 805 810 815
 Asp Asp Asn Ser Gln Ala Pro Leu Arg Leu Gly Ile Asp Met Pro
 820 825 830
 35 Ser Val Val Ile Thr Ser Thr Ser Gln Tyr Phe Thr Leu Tyr Val
 835 840 845
 Ile Ile Val Ser Leu Leu Phe Tyr Ser Glu Pro Met Ser Lys Val Ile
 850 855 860
 His Lys Lys Ile Glu Lys Met Lys Phe Ser Ile Asp Phe Glu Asp Leu
 40 865 870 875 880
 Gly Ala Leu Thr Ser Arg Leu Thr Lys Met Gln Gln His His Lys Leu
 885 890 895
 Leu Lys Val Leu Ser Xaa Xaa Xaa Xaa Phe Pro Xaa Arg Gly Asn
 45 900 905 910

(2) INFORMATION FOR SEQ ID NO: 20:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 626 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: cDNA

EP 0 982 401 A2

(iii) HYPOTHETICAL: NO

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ATTCTTGTGTT	TGTTTGTGAA	TTTTGATCT	CTTGTCTAGA	ATCACTCATT	AATATTTGAT	60
TCAGGGTTTT	GATTTGCTAA	ATAAGGGTC	TATTAGGAGG	ATATTATATA	TAATGTGATG	120
TGGCGAAAAA	AAAAAAACAAG	ATCTACTACT	CTGTTGGATT	TATTTGTGAT	GGCGATTGAA	180
10 GAGAAAACAC	GTCTTTTAA	CCGCGTTTTT	TATTTTTGG	AGAAGCAAAT	TTCAAGCAAA	240
GACTCTTATT	GTGTTGCITT	TGATCCATTC	AAATTTGTGAA	TTACTTTCA	TTAGAACTAT	300
AACTGTTCAT	TATCAATGAC	GTATACATGT	CTGGTTCTG	TTATGTATTG	TAATTTAGT	360
15 TAATTATAAG	CCGTATATTG	GTAGTATTCC	TCTGTACTCA	CAATGGAATT	GGTCTTCAA	420
CAGCAACAAG	TGTTATTTTC	CCTGAATGTA	AAAAATGAAA	GGTAGTGTGTT	ACATATAGTT	480
GGAAATCAAG	CCTCTGAAAT	GAATCACAAT	ATAATAACAA	TTTGTAGTTG	CAGAGAAAAA	540
20 CAATTCAAGT	TGACGGGTAG	TTTTTTTTT	TTCACTGCAT	TTTCAACGA	AAACTAAATA	600
AAATTTCGCT	GATATTGATA	AAAGTAT				626

(2) INFORMATION FOR SEQ ID NO: 21:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 652 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATGGCGTCAA	TTTCTGTTCC	AATTGAAAAA	GGATCATTTC	ACGATGGAGA	TGGATTCAAT	60
35 CAACATCATT	TAGGAGACCC	AGTTATTTC	GGACCTCCCT	ATATTATTA	ATTATTAAAC	120
TTACCCGTCA	CAGCTAATGA	TTCATTTGTC	CAAGACTTGT	TTCAAAGCAG	ATTTACCCCA	180
40 TATGTCAAAT	TTAAAATTGT	AACAGACCCC	GCATCAAATA	TTTGGAGAC	TCATGTCATT	240
AGACAAAGTGG	CTTTGTGGA	ATTGGAATCG	GCCAGTGATA	TGTCAAAGC	TTTAAATGG	300
CATGATTGT	ATTATAAGAC	AAATAGAAGA	GTAACTGTTG	AAAGTGGCAGA	TTTTAATGAT	360
45 TTTCAAAATT	GTATTAATT	CAATCAAGAA	CATGAACGTG	AAATTATGCA	AATCCAACAA	420
GAATTCATTG	CTCAGAAACA	ACAACAACGG	CAACCCAGAC	ATATGGCTCT	TTTAGATGAA	480
50 TTTGAAAGAA	ACCAGCGCGG	TCCTGGATCA	CCCTTGCATC	AAAACCATGA	TCACCACAAAT	540
CCCCACCCAC	AAACACAACA	ACACCATCAT	TTCAATCCTA	ATTTAACAG	ACCTTCAGGT	600
AGATCAAGTC	TTCCAATAGA	TGAAACGTCT	CATTCAAGAA	GACTTTCTT	TG	652

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

EP 0 982 401 A2

(A) LENGTH: 217 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

5 (iii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Ala Ser Ile Ser Val Pro Ile Glu Lys Gly Ser Phe His Asp Gly
1 5 10 15

Asp Gly Phe Asn Gln His His Leu Gly Asp Pro Val Ile Ser Gly Pro
20 25 30

15 Pro Tyr Ile Ile Lys Leu Leu Asn Leu Pro Val Thr Ala Asn Asp Ser
35 40 45

Phe Val Gln Asp Leu Phe Gln Ser Arg Phe Thr Pro Tyr Val Lys Phe
50 55 60

20 Lys Ile Val Thr Asp Pro Ala Ser Asn Ile Leu Glu Thr His Val Ile
65 70 75 80

Arg Gln Val Ala Phe Val Glu Leu Glu Ser Ala Ser Asp Met Ser Lys
85 90 95

25 Ala Leu Lys Trp His Asp Leu Tyr Tyr Lys Thr Asn Arg Arg Val Thr
100 105 110

Val Glu Val Ala Asp Phe Asn Asp Phe Gln Asn Cys Ile Lys Phe Asn
115 120 125

Gln Glu His Glu Arg Glu Ile Met Gln Ile Gln Gln Glu Phe Ile Ala
130 135 140

30 Gln Lys Gln Gln Gln Arg Gln Pro Arg His Met Ala Leu Leu Asp Glu
145 150 155 160

Phe Glu Arg Asn Gln Arg Gly Pro Gly Ser Pro Leu His Gln Asn His
165 170 175

35 Asp His His Asn Pro His Pro Gln Gln Gln His His His Phe Asn
180 185 190

Pro Asn Leu Asn Arg Pro Ser Gly Arg Ser Ser Leu Pro Ile Asp Glu
195 200 205

40 Thr Ser His Ser Arg Arg Leu Ser Phe
210 215

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1513 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

50 (ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1492

EP 0 982 401 A2

(D) OTHER INFORMATION:/note= "N = A or G or C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

5	GTAGTTGTG AAGAAATTGA ACAATCGGA AAACAACAAT ATCAAACGTG TGCCCAATAA	60
	CACTGTATGT ACCTAGATGG ATTACCAAGA TCTACTACAT AAAATAATAA AGGAGTTCCA	120
	CTCACTCAAA GAGTCACAAAC CATGGGATAG CAGTGTGTTG TATGAGACGT TACTACGATC	180
10	AGTATTAAC ACTTTGATCG AACTTTGGG CATAGACAAT CCACCCAGT ATCTACACCT	240
	CACCAAC AATGATAGTA TAGGTGATTT GAAAATAAAA TACTATGGAA ATGCATTAAG	300
	CAAGTCATC AACGGTCATA GCATGTTGCA ATATCTTGAA TCAAAGCATG TATGGATATT	360
	ACAGGCCGTG GTTGAGATTA TTAATACGCG ATCATATAGA ATCAAAGAGT CTTATTCTGC	420
15	TGTTTCAAA GACGTTTCTC ATTATTTGAA AAAACTACTA AAGGAARAGAT ATGAAGCTGA	480
	ATCTAACTA GAGGATTATA TATTGCAGTG CTTGATGTAC GAGACCCAAT TTTACCAAGG	540
	AATTGTTGAT AATGTTTAA CTGCCGATGA CACCGAAAAA TTGGCTAGTT TTTGGGGAC	600
20	ACGACTATCT GAAGAAGATT CGATGTTAG CTATAGGGAT ATAGATTATC CACTAGAGTT	660
	AAACATTAAT AATGAATCTC TTGAAAAGAT ATATAAAATT TTCTTAGGAG TCATTGGCAC	720
	CAAAGATTC GATATCAAGG AGGTTGCCGTC TGCTGTTGTT GGTGTGTATA AACGACACCA	780
25	GAGAATAGAT CATTGAAA AGTTGGATT AGATGAGATT TTGGGAAAGT TTTTCAGAAA	840
	TATATTGCCA CAACTGTTCC AGAGTGTGAC AAATAAGGT TTCCGGGAAT TTCAACAAAGA	900
	GGTAGATGAC CCACCATCGG ACGTGCTAGA CCAGCTAGAT AATATTGTTG ATGACTTTAT	960
	TGCGGTTGGA ATTGAAGGGG TAGATTTGGG CTTTCCGGCT TTGTTCAAGAC ACTACATAAA	1020
30	ATTCAATGAA CAAATTTTC CCACTGTGGT CGAGGATGCT GACCGCGATT TTGTTGCAAG	1080
	AATTAAAGT TTAATTGCTC AAGTCTTGGG GTTTAAAGAC GATGAAAAAT CCTGTGATAT	1140
	CAATCAAGTG GTATCTGAAT TTGTTTCATT ACAAAAGTTG CTACTTAAGA ATAACATATCT	1200
35	TTCACCATCT ACATTATTGA TGGGTGCAAG TACTCACCGAT TACTATAAAA ATTTACAGAT	1260
	CGTAAAAATA ACCTTTGATG GATGGAATGA GAATTCAAG AGGATATTGA AATTGGAGAA	1320
	CAGCGGCTTT TTACAAAGCA AGACATTGCC AAAGTATTAA AAATTATGGT ACTCAAAAAG	1380
40	TATGAAGTTG AATGAATTAT GTAAACGGGT AGATGAATT TATAATGGAG AACTTTGTCG	1440
	AAAAGTTTG GGCATTGTTG CGAGGGTCAC AACCAAAATG TCTATAAATC CNCAAAATG	1500
	GGAGGGTTGC TGA	1513

(2) INFORMATION FOR SEQ ID NO: 24:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 478 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

50 (ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Asp Tyr Gln Asp Leu Leu His Lys Ile Ile Lys Glu Phe His Ser
 1 5 10 15
 5 Leu Lys Glu Phe Lys Pro Trp Asp Ser Ser Val Leu Tyr Glu Thr Leu
 20 25 30
 10 Leu Arg Ser Val Leu Thr Thr Leu Ile Glu Leu Leu Gly Ile Asp Asn
 35 40 45
 15 Pro Pro Ser Tyr Leu His Leu Thr Thr Asn Asn Asp Ser Ile Gly Asp
 50 55 60
 20 Leu Lys Ile Lys Tyr Tyr Gly Asn Ala Leu Ser Lys Ser Ile Asn Gly
 65 70 75 80
 25 His Ser Met Leu Gln Tyr Leu Glu Ser Lys His Val Ser Ile Leu Gln
 85 90 95
 30 Ala Val Val Glu Ile Ile Asn Thr Arg Ser Tyr Arg Ile Lys Glu Ser
 100 105 110
 35 Tyr Ser Ala Val Phe Lys Asp Val Ser His Leu Phe Glu Lys Leu
 115 120 125
 40 20 Lys Glu Arg Tyr Glu Ala Glu Ser Asn Leu Glu Asp Tyr Ile Leu Gln
 130 135 140
 45 Cys Leu Met Tyr Glu Thr Gln Phe Tyr Gln Gly Ile Val Asp Asn Val
 145 150 155 160
 50 25 Leu Thr Ala Asp Asp Thr Glu Lys Leu Ala Ser Phe Leu Gly Thr Arg
 165 170 175
 55 Leu Ser Glu Glu Asp Ser Met Phe Ser Tyr Arg Asp Ile Asp Tyr Pro
 180 185 190
 60 Leu Glu Leu Asn Ile Asn Asn Glu Ser Leu Glu Lys Ile Tyr Lys Ile
 195 200 205
 65 Phe Leu Gly Val Ile Gly Thr Lys Arg Phe Asp Ile Lys Glu Val Ala
 210 215 220
 70 Ser Ala Val Val Gly Val Tyr Lys Arg His Gln Arg Ile Asp His Phe
 225 230 235 240
 75 35 Glu Lys Leu Asp Ser Asp Glu Ile Leu Gly Lys Phe Phe Arg Asn Ile
 245 250 255
 80 Leu Pro Gln Ser Phe Gln Ser Val Thr Asn Lys Val Phe Arg Glu Phe
 260 265 270
 85 His Lys Glu Val Asp Asp Pro Pro Ser Asp Val Leu Asp Gln Leu Asp
 275 280 285
 90 Asn Ile Val Asp Asp Phe Ile Ala Val Gly Ile Glu Gly Val Asp Leu
 290 295 300
 95 Gly Phe Pro Ala Leu Phe Arg His Tyr Ile Lys Phe Met Asn Glu Ile
 305 310 315 320
 100 45 Phe Pro Thr Val Val Glu Asp Ala Asp Arg Asp Phe Val Ala Arg Ile
 325 330 335
 105 Asn Ser Leu Ile Ala Gln Val Leu Glu Phe Lys Asp Asp Glu Lys Ser
 340 345 350
 110 50 Cys Asp Ile Asn Gln Val Val Ser Glu Phe Val Ser Leu Gln Ser Leu
 355 360 365
 115 Leu Leu Lys Asn Asn Tyr Leu Ser Pro Ser Thr Leu Leu Met Arg Ala

EP 0 982 401 A2

370 375 380
Ser Thr His Asp Tyr Tyr Lys Asn Leu Gln Ile Val Lys Ile Thr Phe
385 390 395 400

5 Asp Gly Trp Asn Glu Asn Ser Lys Arg Ile Leu Lys Leu Glu Asn Ser
405 410 415

Gly Phe Leu Gln Ser Lys Thr Leu Pro Lys Tyr Leu Lys Leu Trp Tyr
420 425 430

10 Ser Lys Ser Met Lys Leu Asn Glu Leu Cys Asn Arg Val Asp Glu Phe
435 440 445

Tyr Asn Gly Glu Leu Cys Arg Lys Val Leu Gly Ile Val Gly Arg Val
450 455 460

15 Thr Thr Lys Met Ser Ile Asn Xaa Gln Lys Trp Glu Gly Cys
465 470 475

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:
AGTTATGTCT CATACTACAC ACACAGATGA GGACATGTGT TTAAATGATA AATTGAAATA 60

30 TTTGTACGAT TTATAATCGC TTTATCGTGA CAATTCGAA TACTGGTACT TTCTACTCTA 120

TTTGACAAAA ATTTGCAAAA AATTGGGAA AAAATCCTG TTGCATTTTC GAGACCATCA 180

GTTGCAACCA ATCTGAATAT ATTTTGACAC TTCATAAAAT CTAGTGAAAC TAGCGTCTA 240

CTTTTTAATT CTAATCATCT CATACTATAT CAAGCAAAGA CTTACTATGC GTTTATCAA 300

35 TTTAAGAAAA TGTAGACAGT ACGAAAATAC ACGAGTTCC CAATCTTGA ACTTGAAAAG 360

ATAGTAATAC CGAGATTGGC CAAATCCTAG CCATAGTCCG TTCATACAAA TTCATGAACA 420

ACATCTACAT AAGTAA 436

(2) INFORMATION FOR SEQ ID NO: 26:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:
CTTCTTTCG AATTAGATTG AATCTTTCC AATTTTGCTT GTACACTTGC TAGTTGAAT 60

EP 0 982 401 A2

5	TTACGTTTT CCTCTTTACG TTGTTTCACA ATGGCTGCAC GTTCTCAAA ATTTATTCCC	120
	TTCTTCTTGG TTGGTCTTAT ATCGTTCTCA TCTTCAGGCT TCCTCTCCTC TTGTAACCT	180
	TCTTTTCTA ATAGTTGAA ATAGTTCTT CTTAATCTAG CCCTATGGGT TAATGCACGT	240
	TTTATATCTT GAGACTTGGC TTCTCGACGA TCTATAAATT TCTTTTTGA TTTAAATGAA	300
10	TTTTTATTAT TTGGATGCCAT TTGGTGTGGAG GTGTATTGAA TAGGTGATA ACTAGAAATA	360
	AAAACATGT GAAAGAACAA ATGCCAATC ACTAAAAAAA ATTTAAGATG AGTATGAAAT	420
	CAAAACTTTA CGACATCTT CCGACATGCA CATTATGAGC GACATTTGA TTGATACCA	480
	GAAATAGACA GATTTAGACA GGGTCTATAA CAGAGAAATC AACAAATTAAAC TGGTATCAAC	540
	CTTAAGATTA AAAATGGTCT ATGGCGATAT GAACTGTGT GATGAAAAAC AATATATTTG	600
15	GAAATACCTC TTTTCATTTG ACAATTTTT ATAAAATTTT GGCAACAAATT TTGTACCTAA	660
	AAATTCTTTT GTCTTCAAAAA GTGAAATGTA ATATAGAAAT ACTATTACAA CAAACAA	717

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 667 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:	
30	TTTGTGTTA TATTGATGAT GTTTTTAAGT GCTTGTGTTAT CATGGTGGAT GGAAATTAGA	60
	ATGAGTAAAT TGAATGGAAA ATCACTGCAA CACCAACAAAC AACCAACTGGT GGATACGAAA	120
	ATTTAGTGTAA CAAATTCTG CCAAAAAAAAT ACAATAAAA CCGCTTATAG TCTTCTACTG	180
	ACATAACAAAC ACAAGTCAAT AAATCAACAA CTCATAACAA ATGTAGACCTT AATACTATCG	240
35	CTTAATTATT TAAACTATAA TAAATACCC ATAGTATTAT GCCTTGTCA ATGTGTGTAG	300
	AAATTGGTTA TTACATATCC ATGTGTAAATA TATATGTTGA TCAAAAAACG CGATCTTCTC	360
	TTTGGTGTAG TGTGTTACAC AAAAATTCA CTAGTCTAGG TCACATGATA ATCAGGTGAA	420
	AAATCAAAAT TTGTTGAAT TGAATTTCCT CAATTTGAA ATTTTGTGAA AAATTTTTT	480
40	TTTGCTTAC AAAAAGACTC CATTGTTTT TCCATTCAC AACCAATTAC TTAATTCTC	540
	TTTTTCATAA TTAATAACTA TCATTACTTA CAACTACAAA CAACTACGGT CATTCCCTAA	600
	GAAGGAGCAA CGAGGGCAGA TTGAGACATT AATCCCCTTT ATTTTATCAT CATGCCCTAT	660
45	ACAGAAC	667

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 165 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:
 AACTATTGCC AATGGTAAAT ATGCCAGTGA AATCGAGAAAT TTTAATAAGT CGGTCCCTCT 60
 TAAGGTCCCA TTCAAATTCA CTAATGCACA ATTGGATCTT TATGCTGCTA GCACACATAA 120
 CCAAGAGCCA ATATCCTAGT AACGACGCAC CATACTAGAC CGAAT 165

10 (2) INFORMATION FOR SEQ ID NO: 29:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 207 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO

20 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:120
 (D) OTHER INFORMATION:/note= "N = A or C or G or T"

25 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:129
 (D) OTHER INFORMATION:/note= "N = A or T or C or G"

30 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:162
 (D) OTHER INFORMATION:/note= "N = A or T or C or G"

35 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:178
 (D) OTHER INFORMATION:/note= "N = A or T or C or G"

40 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:194
 (D) OTHER INFORMATION:/note= "N = A or T or C or G"

45 (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:199
 (D) OTHER INFORMATION:/note= "N = A or T or C or G"

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:
 ATGAAGATTT CACCAAGAGAC AGTAAATAAA CTACAACTGG ATGCATCGTG TATAAGAAAC 60

EP 0 982 401 A2

ATCTGTATTT TAGCACATGT CGACCACGGT AAAACCTCAT TGAGTGACTC ATTATTAGCN	120
ACCAATGGNA TCATTTCCA ACGTATGGCA GGTAAGTTA GNTATCTGAA TTGAGANGA	180
GATGACAAAT TGANNGGTNT AACATG	207

5

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 69 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

10

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

Met Lys Ile Ser Pro Glu Thr Val Asn Lys Leu Gln Ser Asp Ala Ser			
1	5	10	15
10	15		

20

Cys Ile Arg Asn Ile Cys Ile Leu Ala His Val Asp His Gly Lys Thr			
20	25	30	35
30	35		

25

Ser Leu Ser Asp Ser Leu Leu Xaa Thr Asn Xaa Ile Ile Ser Gln Arg			
35	40	45	50
45	50		

Met Ala Gly Lys Val Xaa Tyr Leu Asp Ser Arg Xaa Asp Glu Gln Leu			
50	55	60	65
60	65		

(2) INFORMATION FOR SEQ ID NO: 31:

30

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2510 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 2481
 (D) OTHER INFORMATION:/note= "N = A or T or C or G"

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

AAGTCATGCG ATTGCAACAA GGATCACAGA AACCGAGAGT TCACGAACAT TTGATTAATT	60
---	----

45

TGATTGATTC ACCTGGGCAT ATTGACTTTT CGTCTGAAGT GAGTACTTCT TCGAGATTAT	120
---	-----

50

GTGATGGTC AGTTGTTTG GTGGATGTCG TCGAAGGTGT CTGCTCACAA ACAGTCACG	180
--	-----

TTCTACGCCA ATGTTGGATT GATAAGTTGA AGCCATTACT AGTTATTAC AAAATTGATA	240
--	-----

GGTTAATCAC AGAATGGAAA TTGTCTCCCT TGGAGGCATA CCAACACATT TCCAGAATTA	300
---	-----

TAGAACAACT AAACTCTGTG ATTGGTCAT TTTTGCTGG TGATAGACTA GAAGATGACT	360
---	-----

TGAATTGGCG TGAGGCTGGT TCTGTCGGGG AGTTTATCGA GAAGAGTGAT GAAGACTTGT	420
---	-----

55

EP 0 982 401 A2

	ATTCACACC TGAAAAGAAT AATGTAATAT TTGCCTCGGC AATAGATGGA TGGGCATTT	480
5	CAGTCAATAC ATTCGCCAAA ATATACTGAA AAAAATTAGG GTTCTCTCAA CAAGCATTGT	540
	CAAAAACCTCT CTGGGGAGAC TTTTACTTGG ATATGAAAAA TAAAAAAAATC ATCCCTGGTA	600
	AAAATTGAA AAATAATAGT AACAGTTGA AGCCATTATT TGTTTCGTTG ATTTGGACC	660
	AGGTTTGGC TGTTTATGAA AACTGTGTTA TTGAAAGAAA TCAAGACAAG TTGGAAAAAA	720
10	TCATTGAGAA ATTAGGGGCC AAAATCACCC CTCGTGATTG GCGATCCAAA GATTACAAGA	780
	ACTTGCTAAA CTTGATTATG TCTCAGTGGA TTCCCTTGAG TCATGCCATA TTGGGGTCAG	840
	TGATTGAATA CTTGCCAAGC CCCATTGTTG CTCAGCGTGA AAGAATAGAC AAGATTTGG	900
	ATGAAACGAT TTATAGTGCA GTGGATTCAAG AACTGAGATA AATCCAACT AGTCGACCC	960
15	TCATTGTCA AGCGGATGCA GGAATGTGAT AGTCACACC CGGAAACCCA TACAATAGCA	1020
	TATGTATCAA AATTGTTGTC AATCCCCAAT GAAGACTTAC CCAAAGCTAG TAATGCCGCT	1080
	ACTGGAGGAT TGACGGCCGA TGAAATCCAA GAACGAGGAA GAATTGCTCG AGAATTAGCC	1140
	AAAAAGGCAT CTGAAGCAGC TGCTTGGCA CAAGAAGGTT CCAAAATGA AGATGAGTTT	1200
20	GCCATTAAAC CCAAGAAAAGA TCCATTGAA TGGGAATTG AGGAGGACGA TTTTGAGAAT	1260
	GAGGAAGATG AGAGCGATGC AAACCGAGTT GAAGAATCAA CTGAAACCAT AGTGGGTTTC	1320
	ACTCGTATTT ATTCTGGATC GTTATCTAGA GGCCAAAAGC TCACGGTAAT TGGACCCAAA	1380
25	TACGACCCCTT CATTACCTAG AGACCATCAA ACCAACTTTG AACAAATAAC CAATGAAGTT	1440
	GAAATTAAAG ACTTGTTTTT AATCATGGGA CGAGAATTAG TGAGAATGGA AAAAGTCCTG	1500
	CGGGTAATAT TGTGGGGTT GTTGGATTGG ATACGCCGTG CTTAAGAATG CCACAATTG	1560
	CTCACCGTTA CCTGAAGATA AACCACAT TAATTTAGCT TCAACATCAA CTTGATCCA	1620
30	CAATAAACCA ATTATGAAAA TAGCAGTTGA ACCAACAAAC CCAATAAAAC TAGCAAATT	1680
	GGAAACGGAGA TTAGATTAT TGGCCAAAGC CGACCCGGGT TTGGAATGGT ATGTCGACGA	1740
	CGAGTCAGGT GAATTGATTG TTTGTCTTGC TGGAGAATTG CATCTAGAAC GATGCTTGAA	1800
35	AGATTAGAA GAGAGATTGG CTAAGGGTTG TGAAGTTACC GTCAAAGAGC CAGTCATTCC	1860
	CTTCAGAGAG GGGTTGGCAG ATGACAAAT CAGTACCAAC ACCAATAATA ACAACGACGA	1920
	CAATGAAGAT CATGAATTAG ATGAAAACGA AGATGAGCTT GCTGATTAG AGTTGATAT	1980
	TTCTCCGTG CCATTAGAAG TGACTCAGTT TTTAATTGAG AATGAAACGA TTATTGCCGA	2040
40	AATTGTCAAC AACAAAGCAAG ATACTCATGA AATTAGAAAC GATTTTATTG AAAAATTGCA	2100
	CACTATTATT GATAATTCTA ATTTGGCTAC ACAATTCCA GACACCAAGT CTTTTATCAA	2160
	CAATATAATT TGCTTGGAC CTAAACGTGT TGGGCCTAAT ATTTCATGG AAGATTATGG	2220
45	GTAAACAAA TTTAGACATC TACTTGGTGA ATCTGCCACT GAATCTCGAT TTGTTTATGA	2280
	GAATAATGTG TTCAATGGGG TTCAATTGGT ATTCAATGGG GGTCCGTAG CATCAGAGCC	2340
	AATGCAAGGT ATTATTGTTA GACTTAAGAA GCCAGAAAAA AGAGAAGTTG ACGAGGATAA	2400
	GATAGTCACAC CCTGGTAAAAA TAATCACACA GACTCGTGA TTGATTACA AGCGGTTTT	2460
50	GCAAAATCA CCACGCTTGT NCCTTGCAAT GTATACGTGT GAAATCCAAG	2510

(2) INFORMATION FOR SEQ ID NO: 32:

EP 0 982 401 A2

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

5 (ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Val Met Arg Leu Gln Gln Gly Ser Gln Glu Pro Glu Val His Glu His
1 5 10 15

Leu Ile Asn Leu Ile Asp Ser Pro Gly His Ile Asp Phe Ser Ser Glu
15 20 25 30

Val Ser Thr Ser Ser Arg Leu Cys Asp Gly Ala Val Val Leu Val Asp
20 35 40 45

Val Val Glu Gly Val Cys Ser Gln Thr Val Asn Val Leu Arg Gln Cys
20 30 35 40 45 50 55 60

Trp Ile Asp Lys Leu Lys Pro Leu Leu Val Ile Asn Lys Ile Asp Arg
20 30 35 40 45 50 55 60 65 70 75 80

Leu Ile Thr Glu Trp Lys Leu Ser Pro Leu Glu Ala Tyr Gln His Ile
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95

Ser Arg Ile Ile Glu Gln Val Asn Ser Val Ile Gly Ser Phe Phe Ala
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110

Gly Asp Arg Leu Glu Asp Asp Leu Asn Trp Arg Glu Ala Gly Ser Val
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 115 120 125

Gly Glu Phe Ile Glu Lys Ser Asp Asp Leu Tyr Phe Thr Pro Glu
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140

Lys Asn Asn Val Ile Phe Ala Ser Ala Ile Asp Gly Trp Ala Phe Ser
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160

Val Asn Thr Phe Ala Lys Ile Tyr Ser Lys Lys Leu Gly Phe Ser Gln
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175

Gln Ala Leu Ser Lys Thr Leu Trp Gly Asp Phe Tyr Leu Asp Met Lys
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190

Asn Lys Lys Ile Ile Pro Gly Lys Lys Leu Lys Asn Asn Ser Asn Ser
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205

Leu Lys Pro Leu Phe Val Ser Leu Ile Leu Asp Gln Val Trp Ala Val
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220

Tyr Glu Asn Cys Val Ile Glu Arg Asn Gln Asp Lys Leu Glu Lys Ile
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240

Ile Glu Lys Leu Gly Ala Lys Ile Thr Pro Arg Asp Leu Arg Ser Lys
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255

Asp Tyr Lys Asn Leu Leu Asn Leu Ile Met Ser Gln Trp Ile Pro Leu
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270

Ser His Ala Ile Leu Gly Ser Val Ile Glu Tyr Leu Pro Ser Pro Ile
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285

Val Ala Gln Arg Glu Arg Ile Asp Lys Ile Leu Asp Glu Thr Ile Tyr
20 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300

EP 0 982 401 A2

Ser Ala Val Asp Ser Glu
305 310

(2) INFORMATION FOR SEQ ID NO: 33:

5

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

10

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Asp Lys Ser Lys Leu Val Asp Pro Ser Phe Val Lys Ala Met Gln Glu
1 5 10 15

20

Cys Asp Ser Ser His Pro Glu Thr His Thr Ile Ala Tyr Val Ser Lys
20 25 30

25

Leu Leu Ser Ile Pro Asn Glu Asp Leu Pro Lys Ala Ser Asn Ala Ala
35 40 45

Thr Gly Gly Leu Thr Ala Asp Glu Ile Gln Glu Arg Gly Arg Ile Ala
50 55 60

30

Arg Glu Leu Ala Lys Lys Ala Ser Glu Ala Ala Ala Leu Ala Gln Glu
65 70 75 80

Gly Ser Lys Asn Glu Asp Glu Phe Ala Ile Lys Pro Lys Lys Asp Pro
85 90 95

35

Phe Glu Trp Glu Phe Glu Glu Asp Asp Phe Glu Asn Glu Glu Asp Glu
100 105 110

40

Ser Asp Ala Asn Ala Val Glu Glu Ser Thr Glu Thr Ile Val Gly Phe
115 120 125

Thr Arg Ile Tyr Ser Gly Ser Leu Ser Arg Gly Gln Lys Leu Thr Val
130 135 140

Ile Gly Pro Lys Tyr Asp Pro Ser Leu Pro Arg Asp His Gln Thr Asn
145 150 155 160

Phe Glu Gln Ile Thr Asn Glu Val Glu Ile Lys Asp Leu Phe Leu Ile
165 170 175

Met Gly Arg Glu Leu Val Arg Met Glu Lys Val Ser
180 185

(2) INFORMATION FOR SEQ ID NO: 34:

45

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 336 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

50

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

55

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Gly Asn Ile Val Gly Val Val Gly Leu Asp Xaa Ala Val Leu Lys Asn
1 5 10 15

5 Ala Thr Ile Cys Ser Pro Leu Pro Glu Asp Lys Pro Tyr Ile Asn Leu
20 25 30

Ala Ser Thr Ser Thr Leu Ile His Asn Lys Pro Ile Met Lys Ile Ala
35 40 45

10 Val Glu Pro Thr Asn Pro Ile Lys Leu Ala Lys Leu Glu Arg Gly Leu
50 55 60

Asp Leu Leu Ala Lys Ala Asp Pro Val Leu Glu Trp Tyr Val Asp Asp
65 70 75 80

Glu Ser Gly Glu Leu Ile Val Cys Val Ala Gly Glu Leu His Leu Glu
85 90 95

15 Arg Cys Leu Lys Asp Leu Glu Glu Arg Phe Ala Lys Gly Cys Glu Val
100 105 110

Thr Val Lys Glu Pro Val Ile Pro Phe Arg Glu Gly Leu Ala Asp Asp
115 120 125

20 Lys Ile Ser Thr Asn Thr Asn Asn Asn Asp Asp Asn Glu Asp His
130 135 140

Glu Leu Asp Glu Asn Glu Asp Glu Leu Ala Asp Leu Glu Phe Asp Ile
145 150 155 160

25 Ser Pro Leu Pro Leu Glu Val Thr Gln Phe Leu Ile Glu Asn Glu Thr
165 170 175

Ile Ile Ala Glu Ile Val Asn Asn Lys Gln Asp Thr His Glu Ile Arg
180 185 190

Asn Asp Phe Ile Glu Lys Phe Ala Thr Ile Ile Asp Asn Ser Asn Leu
195 200 205

30 Ala Thr Gln Phe Pro Asp Thr Lys Ser Phe Ile Asn Asn Ile Ile Cys
210 215 220

Phe Gly Pro Lys Arg Val Gly Pro Asn Ile Phe Ile Glu Asp Tyr Gly
225 230 235 240

35 Leu Asn Lys Phe Arg His Leu Leu Gly Glu Ser Ala Thr Glu Ser Arg
245 250 255

Phe Val Tyr Glu Asn Asn Val Phe Asn Gly Val Gln Leu Val Phe Asn
260 265 270

40 Gly Gly Pro Leu Ala Ser Glu Pro Met Gln Gly Ile Ile Val Arg Leu
275 280 285

Lys Lys Ala Glu Lys Arg Glu Val Asp Glu Asp Lys Ile Val Asn Pro
290 295 300

45 Gly Lys Ile Ile Thr Gln Thr Arg Asp Leu Ile Tyr Lys Arg Phe Leu
305 310 315 320

Gln Lys Ser Pro Arg Leu Xaa Leu Ala Met Tyr Thr Cys Glu Ile Gln
325 330 335

(2) INFORMATION FOR SEQ ID NO: 35:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 841 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

EP 0 982 401 A2

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 8
(D) OTHER INFORMATION:/note= "N = A or T or C or G"

10

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 9
(D) OTHER INFORMATION:/note= "N = A or T or G or C"

15

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 18
(D) OTHER INFORMATION:/note= "N = A or T or C or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

20

CGCGAAGNN CAATCATNTC AGAAGAAAATG AAAGAAGGTA CTCCGTTCTT TACTATTGTG 60

GCAAGAATCC CTGTGATTGA GGCATTTGGG TTTCCGAGG ATATTAGAAA GAAGACATCC 120

GGGGCAGCTA GTCCCTCAATT AGTTTTGAT GGTATGATA TGTTAGATAT CGATCCATT 180

TGGGTTCCAC ATACTGAAGA AGAATTAGAA GAATTGGGTG AATTTCCAGA AAGAGAAAAT 240

25

GTTGCTAGAA GATATATGAA TAATATCAGA AGAAGAAAAG GTTTATTGT TGATGAGAAA 300

GTCGTCAAAATGCTGAAAAA GCAAAGAACT TTGAAAAGAG ATTAGATTAT CCAGTAAAC 360

AGGCAATATG TGTGAAATTG TTACAGAAAA GACAGATACG ATGTGCCAT TATTTGTTA 420

30

ATATTCAACA ACAAGTAAAT GTATTGATAT AGATGTATAA TATAGTCAAA TGTTGAGACT 480

ATCCGAATAG ACATAGACAC ACAACTCAGC CTGTCAGGGC TGTTTATTAA GTTGTGATGT 540

ATACTAAAAT CCATCCACAC TTCTCGTAAT TGTAGGGAAAG AATTACAAAAA AAGATCACAT 600

AAAAATAATA ATTCTATCAC ACTTTGAAAAA TTTGATTGAA GGTGTTACTA GTATTGTTTC 660

35

AACATTACTC TTTCAAAACA ACGAGATCCA AATACTGCAC AATCTCAAA CGAACGGAGT 720

TACATCACTA TAGTTTCTA TTGTTGTAAG ATCAATACAG ACAAAAAGAA AGTGTAGCAT 780

AAATAATTGA TTGCAATTG CCAAACCTAGA AAACAAAGAG GAAAAAAAGA AAAAAATTTC 840

40

A
(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

55

EP 0 982 401 A2

Arg Glu Xaa Ser Ile Xaa Ser Glu Glu Met Lys Glu Gly Thr Pro Phe
1 5 10 15
Phe Thr Ile Val Ala Arg Ile Pro Val Ile Glu Ala Phe Gly Phe Ser
20 25 30
5 Glu Asp Ile Arg Lys Lys Thr Ser Gly Ala Ala Ser Pro Gln Leu Val
35 40 45
Phe Asp Gly Tyr Asp Met Leu Asp Ile Asp Pro Phe Trp Val Pro His
50 55 60
10 Thr Glu Glu Glu Leu Glu Leu Gly Glu Phe Ala Glu Arg Glu Asn
65 70 75 80
Val Ala Arg Arg Tyr Met Asn Asn Ile Arg Arg Arg Lys Gly Leu Phe
85 90 95
15 Val Asp Glu Lys Val Val Lys Asn Ala Glu Lys Gln Arg Thr Leu Lys
100 105 110
Arg Asp

(2) INFORMATION FOR SEQ ID NO: 37:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 564 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
25 (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

30 AACCTAAAAA TGCTTAAGTT CATCAAATCT GGTAAGTTG CTATTGTTGT AAGAGGTCGT 60
TACGCTGGTA AAAAGTAGT CATTGTGAAA CCACATGATG AAGGTACCAA ATCTCACCA 120
TTCCCACATG CCATTGTCGC TGGTATTGAA AGAGCTCCAT TGAAGGTTAC CAAGAAGATG 180
35 GATGCTAAAAA AAGTTACCAA AAGAACTAAA GTCAAGCCAT TTGTTAAATT AGTAAACTAC 240
AACCATTAA TGCCAACTAG ATACTCATTG GATGTTGAAT CATTCAAATC TGCTGTCACT 300
TCTGAAGCTT TAGAAGAAC ATCTCAAAGA GAAGAAGCTA AAAAGTTGT CAAGAAGGCT 360
40 TTTGAAGAAA AACATCAAGC TGGTAAGAAC AAATGGTTCT TCCAAAAATT ACACTTTAA 420
GAAAGGAACC ACCTTTATTT GAATGTTGT AATATAGGTT GAATCAGAGA GACAAAGTAG 480
AAGAAAATAC AAAAAAGAGA GTATATCTGT ATAGTATAAT TTAATGGGGG TCTAATTTAC 540
TTACCACTTT ATTCTGTCAT TATT 564

45 (2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 136 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown
50 (ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

5 Met Ala Lys Phe Ile Lys Ser Gly Lys Val Ala Ile Val Val Arg Gly
1 5 10 15
Arg Tyr Ala Gly Lys Lys Val Val Ile Val Lys Pro His Asp Glu Gly
5 20 25 30
Thr Lys Ser His Pro Phe Pro His Ala Ile Val Ala Gly Ile Glu Arg
35 40 45
10 Ala Pro Leu Lys Val Thr Lys Lys Met Asp Ala Lys Lys Val Thr Lys
50 55 60
Arg Thr Lys Val Lys Pro Phe Val Lys Leu Val Asn Tyr Asn His Leu
65 70 75 80
15 Met Pro Thr Arg Tyr Ser Leu Asp Val Glu Ser Phe Lys Ser Ala Val
85 90 95
Thr Ser Glu Ala Leu Glu Glu Pro Ser Gln Arg Glu Glu Ala Lys Lys
100 105 110
20 Val Val Lys Lys Ala Phe Glu Glu Lys His Gln Ala Gly Lys Asn Lys
115 120 125
Trp Phe Phe Gln Lys Leu His Phe
130 135

(2) INFORMATION FOR SEQ ID NO: 39:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1192 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
30 (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

TTTGAAACGA TTAAGTCCAA TCAAACAATC TTATTCAAAA GTACTCGCAA TACGTACAAT 60
GTCAATTCCA TCTACTCAGT ACGGATTTTT TTATAATAAA GCTAGTGGTC TTARTTGAA 120
40 AAAAGACTTG CCGGTTAACCA AGCCAGGTGC TGGTCAATTG CTTTTAAAGG TTGATGCAGT 180
TGGCCTTTGT CATTCAAGATT TACATGTTCT CTATGAAGGT TTGGATTGTG GTGATAATT 240
TGTGATGGGC CACGAAATTG CTGGGACTGT TGCTGAACTA GGTGAAGAGG TGAGTGAGTT 300
TGCAGTTGGA GATCGTGTGCG CTTGTGTCGG CCCCCAATGGA TGTGGCTTT GTAAACACTG 360
45 TCTTACTGGT AACGATAATG TTTGTACCAA GTCGTTTTG GATTGGTTG GATTGGGTTA 420
CAATGGAGGT TACGAGCAAT TTTTGTAGT CAAGAGACCA AGAAACTTGG TCAAGATCCC 480
TGACAATGTT ACTTCCGAGG AAGCTGCAGC TATTACGGAT GCCGTATTGA CTCCTTACCA 540
50 TGCTATCAAG TCTGCAGGTG TTGGTCCAGC AAGTAATATA TTAATTATCG GAGCTGGTGG 600
ATTAGGAGGT AACGCTATTTC AAGTTGCAAA AGCATTGGT GCGAAGGTTA CTGTTTGGA 660

EP 0 982 401 A2

	TAAAAAGGAT AAGGCAAGAG ACCAAGCTAA GGCCTTGGA CCTGACCAGG TTTACAGTGA	720
	ATTACCAAGAC AGCGTTTAC CTGGGTCAATT CAGTGCTTGT TTTGATTTG TTTCCGGTCA	780
5	GGCAACATAC GATTGTGTC AAAAGTATTG TGAGCCAAG GGTACTATGG TTCCCGTAGG	840
	TCTAGGTGCA ACTTCGCTTA ACATAAAATCT TGCTGATTAA GATCTTCGTG AAATTACCGT	900
	CAAGGGCTCA TTCTGGGTA CCCTGATGGA TTAAAGAGAA GCATTTGAAT TGGCTGCACA	960
10	GGGAAAGGTC AAACCAAATG TTGCTCATGC TCCATTGTCA GAATTGCCTA AGTATATGGA	1020
	GAAGTTGAGA GCCGGTGGTT ATGAAGGAAG AGTCGTGTTT AATCCATAAT ACTGAAAAGT	1080
	GAAGAAACCA TCAATAATAG CTTGGTGAGT ATGTATGGGA AATATTCAATT TATGTATGTA	1140
	GGTCATTTAT ATGTGTGAA TGATTTCTAA TCTGAATTTC GTACAATTCT TT	1192

(2) INFORMATION FOR SEQ ID NO: 40:-

15

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 336 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

20

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

Met	Ser	Ile	Pro	Ser	Thr	Gln	Tyr	Gly	Phe	Phe	Tyr	Asn	Lys	Ala	Ser
1															15

Gly	Leu	Asn	Leu	Lys	Lys	Asp	Leu	Pro	Val	Asn	Lys	Pro	Gly	Ala	Gly
20															30

30

Gln	Leu	Leu	Leu	Lys	Val	Asp	Ala	Val	Gly	Leu	Cys	His	Ser	Asp	Leu
35															45

His	Val	Leu	Tyr	Glu	Gly	Leu	Asp	Cys	Gly	Asp	Asn	Tyr	Val	Met	Gly
50															60

35

His	Glu	Ile	Ala	Gly	Thr	Val	Ala	Glu	Leu	Gly	Glu	Val	Ser	Glu	
65															80

Phe	Ala	Val	Gly	Asp	Arg	Val	Ala	Cys	Val	Gly	Pro	Asn	Gly	Cys	Gly
85															95

40

Leu	Cys	Lys	His	Cys	Leu	Thr	Gly	Asn	Asp	Asn	Val	Cys	Thr	Lys	Ser
100															110

Phe	Leu	Asp	Trp	Phe	Gly	Leu	Gly	Tyr	Asn	Gly	Gly	Tyr	Glu	Gln	Phe
115															125

Leu	Leu	Val	Lys	Arg	Pro	Arg	Asn	Leu	Val	Lys	Ile	Pro	Asp	Asn	Val
130															140

45

Thr	Ser	Glu	Glu	Ala	Ala	Ala	Ile	Thr	Asp	Ala	Val	Leu	Thr	Pro	Tyr
145															160

His	Ala	Ile	Lys	Ser	Ala	Gly	Val	Gly	Pro	Ala	Ser	Asn	Ile	Leu	Ile
165															175

50

Ile	Gly	Ala	Gly	Gly	Leu	Gly	Gly	Asn	Ala	Ile	Gln	Val	Ala	Lys	Ala
180															190

Phe	Gly	Ala	Lys	Val	Thr	Val	Leu	Asp	Lys	Lys	Asp	Ala	Arg	Asp
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

55

EP 0 982 401 A2

	195	200	205	
	Gln Ala Lys Ala Phe Gly Ala Asp Gln Val Tyr Ser Glu Leu Pro Asp			
5	210	215	220	
	Ser Val Leu Pro Gly Ser Phe Ser Ala Cys Phe Asp Phe Val Ser Val			
	225	230	235	240
	Gln Ala Thr Tyr Asp Leu Cys Gln Lys Tyr Cys Glu Pro Lys Gly Thr			
10	245	250	255	
	Ile Val Pro Val Gly Leu Gly Ala Thr Ser Leu Asn Ile Asn Leu Ala			
	260	265	270	
	Asp Leu Asp Leu Arg Glu Ile Thr Val Lys Gly Ser Phe Trp Gly Thr			
	275	280	285	
	Ser Met Asp Leu Arg Glu Ala Phe Glu Leu Ala Ala Gln Gly Lys Val			
15	290	295	300	
	Lys Pro Asn Val Ala His Ala Pro Leu Ser Glu Leu Pro Lys Tyr Met			
	305	310	315	320
	Glu Lys Leu Arg Ala Gly Gly Tyr Glu Gly Arg Val Val Phe Asn Pro			
	325	330	335	

20 (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2021 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1270
- (D) OTHER INFORMATION:/note= "R = A or G"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1395
- (D) OTHER INFORMATION:/note= "R = A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

40	ATGGAAAAAA TTGACATTAA TACAAATTCA AACAAAATCC AACAAAGCATA CGATAAAGTT	60
	GTTAGAGGAG ACCCAAATGC AACATTCGTC GTTATTCTG TTGACAAAAA CGCCACTATG	120
	GACGTCACTG AAACAGGGGA CGGATCATTA GAGGATTTG TTGACACATT TACTGATGGA	180
	CAAGTTCAAT TTGCTTTAGC CAGGGTTACT GTTCCAGGAT CTGACGTTTC CAAAACATC	240
45	TTGTTAGGAT GGTGTCCTGA CAGTGCTCCA GCAAAATTGA GATTGTCATT TGCCAATAAT	300
	TTTGCTGATG TGTCCAGAGT ATTGAGCGGA TACCATGTCG AAATTACTGC AAGGGATCAA	360
	GATGATTTAG ACGTGAATGA ATTCTTGAAT AGAGTTGGTG CTGCTGCTGG TGCAAGATAT	420
50	TCCACTCAAA CTTCCGGACT CAAAAAACCA TCCCCTGCTG CACCTAAACC TACTTCAAAA	480
	CCTGTTGTTG CTAATCTAG TTCTGCTTC AAACCTTCAT TTGTACCCAA ATCTACTGGG	540
	AAGCCTGTTG CTCCAGCTAA GCCAAAACCA AAGAACATCA CCAAGGATGC TGGTTGGGT	600

EP 0 982 401 A2

	GATGCTGAAG ACGTTGAGGA AAGAGACTTT GACAAGAAC CTTGGATAA CGTTCATCG	660
5	GCATATAAAC CAACAAAGGT TAACATTGAC GAATTGAGAA AACAAAAATC AGATACAAC	720
	AGCTCAACTC CTAAAACATT CAAATCTGAA CCACAGAAG AAAAGAATGA CGATGATGGG	780
	CAATCCAAAC CTTTATCGGA AAGGATGAAA GCCTATGATC AACCATCAAG TAGTGATGGA	840
10	AGATTGACTT CTTTACCAAA ACCAAAGATT GGACATTCTG TTGCCGATAA ATATAAGCT	900
	AGTGCATCTG GGAATGGTGC TGCTCTGCG TTTGGTGCTA AACCAAGCATT TGGTACACAA	960
	TCAGTTGATT CAAGAAAGGA TAAATTGGTA GGTGGTTGT CGAGAGATTT TGGTGCTGAA	1020
	AATGGAAAAA CTCCGGCACCA AATTGGGCT GAAAAAAGGG GAAAATACAA AACAGTGGCC	1080
	TCCGATGAGA AAGAAACTAA CTCAAGTGAA AAAGTTGATG ACCCAGAGGA ACATCATGCT	1140
15	CCCGACTTGG CCAAAAAATT TGAAGAAAG GCAAATATTG CTGGCGATAC TCCTTCCTTG	1200
	CCAACTAGAA ACTTACCAACC AGCACCAACCA GCACCGAGAAA CCGCAATTCC ATCTAACGAA	1260
	AAAGACAAAR AAGAAAAGGA AGAGGAAGAA CAAGCTCCAG CACCATCTT GCCTACTAGA	1320
	AACTTACCAAC CACCGTCACA AAGACAACCT GAGCCCGAAC CAGAACCCAGA AGAAGAGGAG	1380
20	GAAGAAGAAG AAGARGAGGC TCCTGCTCCA AGCTTACCAAG CAAAGAAATCT CCCACCCAGCA	1440
	CCAAAAGCAG AAGCAGAAGA ATCAAAAAAA CAGTCACCA CAGCCACCGC AGAGTATGAT	1500
	TACGAAAAGG ACGRAGATAA TGAAATTGGA TTCTCCGAAAG GTGACTTGAT TATTGATATT	1560
25	GAATTTGTGG ATGACGATG GTGGCAAGGT AAACATGCTA AAAACTGGTGA AGTTGGTTTG	1620
	TTTCCTGCCA CTTATGTGTC ATTAATGAA AAAGCTGCTG ACAAAAGAAGA GGAAGCCCCA	1680
	GCTCCAGCTC CAGGCCCATC ATTACCTTCT AGAGAAGAAA CACAAGCAGC ACCAGCATT	1740
	CCAAGTAGAT CAGAGCAAAA ACCAGAACATCA AAAACTGCTA CAGCTGAATA CGATTACGAA	1800
30	AAGGACGAAG ACAATGAAAT TGGTTTTCA GAAGGTGATT TGATTGTTGA AATCGAATTT	1860
	GTGACGATG ATTGGTGGCA AGGAAAACAT TCCAAGACAG GAGAAGTCGG ATTGTTCCCT	1920
	GCTAACTATG TTGTCTGAA TGAGTAGATT TAGTATAAAC AATATTGTT TTTTTTTAT	1980
35	ATGAATCTAT AATATAAAATA CAAAGAAAAG ATAAATTGGT G	2021

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 648 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

Met	Glu	Lys	Ile	Asp	Ile	Asn	Thr	Asn	Ser	Asn	Lys	Ile	Gln	Gln	Ala			
1															15			
5															10			
Tyr	Asp	Lys	Val	Val	Arg	Gly	Asp	Pro	Asn	Ala	Thr	Phe	Val	Val	Tyr			
															30			
20															25			

EP 0 982 401 A2

Ser Val Asp Lys Asn Ala Thr Met Asp Val Thr Glu Thr Gly Asp Gly
 35 40 45
 Ser Leu Glu Asp Phe Val Glu His Phe Thr Asp Gly Gln Val Gln Phe
 50 55 60
 Gly Leu Ala Arg Val Thr Val Pro Gly Ser Asp Val Ser Lys Asn Ile
 65 70 75 80
 Leu Leu Gly Trp Cys Pro Asp Ser Ala Pro Ala Lys Leu Arg Leu Ser
 85 90 95
 Phe Ala Asn Asn Phe Ala Asp Val Ser Arg Val Leu Ser Gly Tyr His
 100 105 110
 Val Gln Ile Thr Ala Arg Asp Gln Asp Asp Leu Asp Val Asn Glu Phe
 115 120 125
 Leu Asn Arg Val Gly Ala Ala Ala Gly Ala Arg Tyr Ser Thr Gln Thr
 130 135 140
 Ser Gly Leu Lys Lys Pro Ser Pro Ala Ala Pro Lys Pro Thr Ser Lys
 145 150 155 160
 Pro Val Val Ala Lys Ser Ser Ala Ser Lys Pro Ser Phe Val Pro
 165 170 175
 Lys Ser Thr Gly Lys Pro Val Ala Pro Ala Lys Pro Lys Pro Lys Asn
 180 185 190
 Ile Thr Lys Asp Ala Gly Trp Gly Asp Ala Glu Asp Val Glu Glu Arg
 195 200 205
 Asp Phe Asp Lys Lys Pro Leu Asp Asn Val Pro Ser Ala Tyr Lys Pro
 210 215 220
 Thr Lys Val Asn Ile Asp Glu Leu Arg Lys Gln Lys Ser Asp Thr Thr
 225 230 235 240
 Ser Ser Thr Pro Lys Thr Phe Lys Ser Glu Pro Gln Glu Glu Lys Asn
 245 250 255
 Asp Asp Asp Gly Gln Ser Lys Pro Leu Ser Glu Arg Met Lys Ala Tyr
 260 265 270
 Asp Gln Pro Ser Ser Asp Gly Arg Leu Thr Ser Leu Pro Lys Pro
 275 280 285
 Lys Ile Gly His Ser Val Ala Asp Tyr Lys Ala Ser Ala Ser Gly
 290 295 300
 Asn Gly Ala Ala Pro Ala Phe Gly Ala Lys Pro Ala Phe Gly Thr Gln
 305 310 315 320
 Ser Val Asp Ser Arg Lys Asp Lys Leu Val Gly Gly Leu Ser Arg Asp
 325 330 335
 Phe Gly Ala Glu Asn Gly Lys Thr Pro Ala Gln Ile Trp Ala Glu Lys
 340 345 350
 Arg Gly Lys Tyr Lys Thr Val Ala Ser Asp Glu Lys Glu Thr Asn Ser
 355 360 365
 Ser Glu Lys Val Asp Glu Pro Glu Glu His His Ala Ala Asp Leu Ala
 370 375 380
 Lys Lys Phe Glu Glu Lys Ala Asn Ile Ala Gly Asp Thr Pro Ser Leu
 385 390 395 400
 Pro Thr Arg Asn Leu Pro Pro Ala Pro Pro Ala Arg Glu Thr Ala Ile
 405 410 415

EP 0 982 401 A2

Pro Ser Asn Glu Lys Asp Lys Xaa Glu Lys Glu Glu Glu Gln Ala
 420 425 430
 Pro Ala Pro Ser Leu Pro Thr Arg Asn Leu Pro Pro Pro Ser Gln Arg
 435 440 445
 5 Gln Pro Glu Pro Glu Pro Glu Glu Glu Glu Glu Glu Glu Glu
 450 455 460
 Xaa Glu Ala Pro Ala Pro Ser Leu Pro Ala Arg Asn Leu Pro Pro Ala
 465 470 475 480
 10 Pro Lys Ala Glu Ala Glu Glu Ser Lys Lys Gln Ser Thr Thr Ala Thr
 485 490 495
 Ala Glu Tyr Asp Tyr Glu Lys Asp Glu Asp Asn Glu Ile Gly Phe Ser
 500 505 510
 15 Glu Gly Asp Leu Ile Ile Asp Ile Glu Phe Val Asp Asp Asp Trp Trp
 515 520 525
 Gln Gly Lys His Ala Lys Thr Gly Glu Val Gly Leu Phe Pro Ala Thr
 530 535 540
 20 Tyr Val Ser Leu Asn Glu Lys Ala Ala Asp Lys Glu Glu Glu Ala Pro
 545 550 555 560
 Ala Pro Ala Pro Ala Pro Ser Leu Pro Ser Arg Glu Glu Thr Gln Ala
 565 570 575
 Ala Pro Ala Leu Pro Ser Arg Ser Glu Gln Lys Pro Glu Ser Lys Thr
 580 585 590
 25 Ala Thr Ala Glu Tyr Asp Tyr Glu Lys Asp Glu Asp Asn Glu Ile Gly
 595 600 605
 Phe Ser Glu Gly Asp Leu Ile Val Glu Ile Glu Phe Val Asp Asp Asp
 610 615 620
 30 Trp Trp Gln Gly Lys His Ser Lys Thr Gly Glu Val Gly Leu Phe Pro
 625 630 635 640
 Ala Asn Tyr Val Val Leu Asn Glu
 645

25 (2) INFORMATION FOR SEQ ID NO: 43:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1340 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 40 (ii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGTGTGACG TCGTATTAGG ATCTCAATGG GGGGATGAAG GTAAAGGTAATTAGTCGAT 60
 TTATTATGTG ATGATATCGA TGTTTGTGCC AGGTGTCAAG GTGGTAACAA TGCTGGCCAC 120
 50 ACAATTGTTG TTGGTAAAGT CAAGTATGAC TTCCACATGT TACCTCTGG TTTGGTCAAT 180
 CCTAAATGTC AAAACTTAGT TGGATCTGGT GTTGTATCC ACGGTCCCTTC CTTCTTGCT 240
 GAATTGGAAA ACTTGGAAAGC AAAAGGGTTA GATTGTCGTG ATAGATTGTT TGTTTCATCT 300

EP 0 982 401 A2

	AGAGCTCATT TGGCTTTGA CTTCCATCAA CGTACTGATA AATTGAAAGA AGCTGAATTA	360
	TCAACCAATA AGAAATCAAT AGGTACTACC GGTAAAGGTA TTGGTCCAAC TTACTCAACC	420
5	AAGGCAAGTA GATCAGGTAT CAGAGTCCAC CATTAGTCACCCGTATCC AGAAGCTTGG	480
	GAAGAATTCA AAACTAGATA TTTGAGATTA GTCGAGAGTA GACAAAAAG ATACGGTGAA	540
	TTGAAATATG ATCCTAAGGA AGAATTGGCA AGATTTGAA AATACCGTGA AACCTTGAGA	600
10	CCATTCGTCG TCGACTCCGT CAACTTCATG CACGAGCTA TTGCTGCCAA TAAAAAAATC	660
	TTGGTTGAAAG TGCTTAATGC GTTAATGTTG GATATTGATT TCGGTACTTA TCCATACGTC	720
	ACTTCTTCAT CAACTGGTAT TGGTGGTGT TTGACTGGGT TGGGTATTCC TCCAAGAACCC	780
15	ATCAGAAATG TCTATGGTGT TGTTAAAGCC TACACCACTA GAGTTGGTGA GGGTCCATTG	840
	CCACACAGAAC AATTGAACAA GGTAGGTGAA ACTTTGCAAG ATGTTGGTGC CGAATATGGT	900
	GTTACTACTG GAAGAAAAAG AAGATGTGGT TGGTTGGATT TGGTTGTGTT GAAATATTCC	960
	AACCTGATCA ACGGATACAC TTCTTGAAAC ATCACCAAAT TGGATGTTT GGATAAAATTC	1020
20	AAGGAAATTG AAGTTGGTGT TGCTTATAAA TTGAATGGAA AAGAGTTGCC AAGTTTCCCT	1080
	GAAGATTGTA TTGATTTAGC TAAAGTCGAG GTTGTGTATA AGAAATTCCC AGCTTGGGAA	1140
	CAAGATATCA CCGGTATCAA GAAATATGAA GACTTGCCAG AAAACGCTAA GAACTATCTT	1200
25	AAATTCAATTG AAGATTACTT GCAAGTTCCA ATCCAATGGG TAGGTACCGG TCCAGCTAGA	1260
	GATTCTATGT TAGAAAAAGA GATTTAGTTG TACACATGCT ACGGAAGACG ATTAGATTG	1320
	TTTTATTAGA TTAATAACCT	1340

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 428 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Met	Cys	Asp	Val	Val	Leu	Gly	Ser	Gln	Trp	Gly	Asp	Glu	Gly	Lys	Gly
1					5					10				15	
Lys	Leu	Val	Asp	Leu	Leu	Cys	Asp	Asp	Ile	Asp	Val	Cys	Ala	Arg	Cys
					20				25			30			
Gln	Gly	Gly	Asn	Asn	Ala	Gly	His	Thr	Ile	Val	Val	Gly	Lys	Val	Lys
					35				40			45			
Tyr	Asp	Phe	Mis	Met	Leu	Pro	Ser	Gly	Leu	Val	Asn	Pro	Lys	Cys	Gln
					50				55		60				
Asn	Leu	Val	Gly	Ser	Gly	Val	Val	Ile	His	Val	Pro	Ser	Phe	Phe	Ala
					65			70		75		80			
Glu	Leu	Glu	Asn	Leu	Glu	Ala	Lys	Gly	Leu	Asp	Cys	Arg	Asp	Arg	Leu
					85				90			95			
Phe	Val	Ser	Ser	Arg	Ala	His	Leu	Val	Phe	Asp	Phe	His	Gln	Arg	Thr

EP 0 982 401 A2

	100	105	110	
	Asp Lys Leu Lys Glu Ala Glu Leu Ser Thr Asn Lys Lys Ser Ile Gly			
	115	120	125	
5	Thr Thr Gly Lys Gly Ile Gly Pro Thr Tyr Ser Thr Lys Ala Ser Arg			
	130	135	140	
	Ser Gly Ile Arg Val His His Leu Val Asn Pro Asp Pro Glu Ala Trp			
	145	150	155	160
10	Glu Glu Phe Lys Thr Arg Tyr Leu Arg Leu Val Glu Ser Arg Gln Lys			
	165	170	175	
	Arg Tyr Gly Glu Phe Glu Tyr Asp Pro Lys Glu Glu Leu Ala Arg Phe			
	180	185	190	
	Glu Lys Tyr Arg Glu Thr Leu Arg Pro Phe Val Val Asp Ser Val Asn			
	195	200	205	
15	Phe Met His Glu Ala Ile Ala Ala Asn Lys Lys Ile Leu Val Glu Gly			
	210	215	220	
	Ala Asn Ala Leu Met Leu Asp Ile Asp Phe Gly Thr Tyr Pro Tyr Val			
	225	230	235	240
20	Thr Ser Ser Ser Thr Gly Ile Gly Gly Val Leu Thr Gly Leu Gly Ile			
	245	250	255	
	Pro Pro Arg Thr Ile Arg Asn Val Tyr Gly Val Val Lys Ala Tyr Thr			
	260	265	270	
25	Thr Arg Val Gly Glu Gly Pro Phe Pro Thr Glu Gln Leu Asn Lys Val			
	275	280	285	
	Gly Glu Thr Leu Gln Asp Val Gly Ala Glu Tyr Gly Val Thr Thr Gly			
	290	295	300	
30	Arg Lys Arg Arg Cys Gly Trp Leu Asp Leu Val Val Leu Lys Tyr Ser			
	305	310	315	320
	Asn Ser Ile Asn Gly Tyr Thr Ser Leu Asn Ile Thr Lys Leu Asp Val			
	325	330	335	
	Leu Asp Lys Phe Lys Glu Ile Glu Val Gly Val Ala Tyr Lys Leu Asn			
	340	345	350	
35	Gly Lys Glu Leu Pro Ser Phe Pro Glu Asp Leu Ile Asp Leu Ala Lys			
	355	360	365	
	Val Glu Val Val Tyr Lys Phe Pro Gly Trp Glu Gln Asp Ile Thr			
	370	375	380	
40	Gly Ile Lys Lys Tyr Glu Asp Leu Pro Glu Asn Ala Lys Asn Tyr Leu			
	385	390	395	400
	Lys Phe Ile Glu Asp Tyr Leu Gln Val Pro Ile Gln Trp Val Gly Thr			
	405	410	415	
45	Gly Pro Ala Arg Asp Ser Met Leu Glu Lys Lys Ile			
	420	425		

(2) INFORMATION FOR SEQ ID NO: 45:

6 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:	
	ATGACTGGTG AAGAAGATAA AAAACAACAT TTTGATGCTT CTGGTGCTTC TGCTGTAGAT	60
	GATAAAAACAG CAACTGCAAT TTTAAGAAGA AAAAAGAAG ATAATGCCTT GGTGCGTTGAT	120
10	GACGCCACCA ACGATGACAA TTCTGTCTA ACCATGTGCGT CAAACACAAT GGAATTGTTA	180
	CAATTATTCC GTGGTGATAC AGTCTTGGTG AAAGGTAAGA AGAGAAAAGGA CACAGTGTG	240
	ATCGTTTAG CTGATGATGA TATGCTGATG GGCCTTGCTA GACTTAACAG ATGTGTTGCT	300
15	AACAATTGC GTGTCAGATT GGGAGATATC GTTACTGTCC ATCCATGTCC TGATATTAAA	360
	TATGCCAACA GAATCTCAGT ATTGCCAATT GCTGATACTG TTGAAGGTAT TAATGGTTCC	420
	TTATTCGACC TTTACTTGAA GCCATATTTT GTTGAAGCCT ATAGACCACT GAGAAAAGGT	480
	GATTATTCA CTGTGAGGGG TGGTATGAGA CAAGTAGAAAT TCAAAGTTGT TGAAGTTGAC	540
20	CCTGAAGAAA TTGCAATTGT TGCTCAAGAT ACCATTATTTC ATTGTGAAGG AGAACCTATT	600
	AATCGTGAAG ATGAAGAAAA TAGCTTGAAT GAACTGGTTT ACGACGATAT TGGAGGTTGT	660
	AAGAAAACAA TGCCCCAAAT TAGAGAATTG GTTGAATTGC CTTTAAGACA TCCACAATTA	720
	TTCAAAATCGA TTGGTATTAA CCCACCAAAAG GGTATTTGA TGTATGGTCC ACCTGGTACC	780
25	GGTAAAACCA TTATGGCAAG AGCAGTGGCC AATGAACAG GTGCCCTCTT TTTCTTAATA	840
	AATGGTCCAG AAAATTATGTC TAAAATGGCT GGTGAGTCTG AATCCAATT AAGAAAAGCT	900
	TTTGAAGAGG CTGAAAAGAA TTCTCCTTCC ATTATTTCA TTGATGAGAT TGACTCTATT	960
	GCCCCAAAGA GAGACAAAAC TAATGGTGAAT CTAGAAAGAA GAGTTGTTTC TCAATTGTTA	1020
30	ACCCCTTATGG ATGGTATGAA GGCCAGATCT AATGTAGTTG TTATTGCTGC TACTAACAGA	1080
	CCAAATTCTA TTGATCCTGC TTTGAGAAGA TTTGGAAAGAT TCGACAGAGA AGTTGACATT	1140
	GGTGTTCGGG ATGCTGAAGG ACGTTAGAG ATTTGAGAA TCCACACAAA GAATATGAAA	1200
35	TTGGCTGATG ATGTTGACTT GGAAGCCATC GCTTCTGAAA CACATGGTTT CGTTGGTGCT	1260
	GATATTGCTT CATTATGTTC AGAAGCTGCT ATGCAACAAA TCCGTGAAAAA GATGGATCTT	1320
	ATCGACTTGG AAGAAGAAC CATTGATACT GAAGTGTGAA ACTCTTTGGG TGTCACTCAA	1380
	GACAACCTCA GATTGCTCT CGGAAACTCC AACCCATCTG CCTTGGTGA AACTGTTGTT	1440
40	GAAAATGTTA ATGTCACCTG GGATGATATT GGTGGTTGG ACAACATTAA GAATGAATTA	1500
	AAAGAAAACCG TGGAGTATCC TGTGTTACAT CCAGATCAAT ACCAAAAAATT CGGATTGGCA	1560
	CCAACAAAAG GTGTTTGTT CTTTGGTCCA CCAGGTACTG GTAAAGACACT TTTGGCCAAG	1620
45	GCTGTTGCTA CTGAAGTTTC TGCTAATTTTC ATTTCTGTCA AAGGTCCAGA ATTGTTGAGT	1680
	ATGTGGTATG GTGAATCTGA GTCTAATATC CGTGATATAT TTGACAAGGC CAGAGCTGCT	1740
	GCTCCTACTG TGGTGGTTT GGATGAATTG GACTCCATTG CCAAAGCTAG AGGTGGTTCT	1800
	CACGGTGATG CTGGTGGTGC CTCCGACAGA GTGGTCAATC AATTGTTGAC TGAATGGAC	1860
50	GGTATGAATG CTAAGAAGAA TGTGTTGTC ATTGGTGCCTA CTAACAGACC AGATCAAATT	1920
	GATCCTGCAT TATTGAGACC AGGTAGATTG GATCAATTAA TTTATGTCCC ATTGCCAGAT	1980

EP 0 982 401 A2

5 GAGCCAGCTA GATTGTCTAT TTTACAAGCT CAATTGAGAA ACACCTCATT AGAACCTGGT 2040
 TTGGACTTGA ACGAAATTGC CAAGATCACT CACGGTTCT CGGGTCCAGA TTTGTCTTAT 2100
 ATTGTTCAAA GATCTGCTAA ATTTGCTATT AAAGACTCTA TTGAAGCCCA AGTAAAGATT 2160
 AACAAAGATTA AAGAAGAGAAA AGAAAAGGTG AAAACTGAAG ATGTTGATAT GAAGGTAGAT 2220
 GAAGTTGAAG AAGAAGACCC TGTGCCTTAC ATTACCAAGAG CTCACCTTGA AGAGGCTATG 2280
 10 AAGACCGCAA AAAGATCTGT TTCAGACGCT GAATTACGTC GTTATGAGTC TTACGCTCAA 2340
 CAATTGCAAG CCTCAAGAGG TCAATTTCT ACCTTTAGAT TCAATGAAA TGCTGGTGCC 2400
 ACTGATAATG GTTCAGGCAGC AGGTGCCAAC TCAGGTGCAG CTTTCCGAAA CGTTGAAGAG 2460
 GAAGACGATT TGTACAGTTG A 2481

15 (2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 826 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

20 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

Met Thr Gly Glu Glu Asp Lys Lys Gln His Phe Asp Ala Ser Gly Ala
 1 5 10 15
 Ser Ala Val Asp Asp Lys Thr Ala Thr Ala Ile Leu Arg Arg Lys Lys
 20 25 30
 Lys Asp Asn Ala Leu Val Val Asp Asp Ala Thr Asn Asp Asp Asn Ser
 35 40 45
 Val Ile Thr Met Ser Ser Asn Thr Met Glu Leu Leu Gln Leu Phe Arg
 50 55 60
 Gly Asp Thr Val Leu Val Lys Gly Lys Lys Arg Lys Asp Thr Val Leu
 65 70 75 80
 Ile Val Leu Ala Asp Asp Asp Met Pro Asp Gly Val Ala Arg Val Asn
 85 90 95
 Arg Cys Val Arg Asn Asn Leu Arg Val Arg Leu Gly Asp Ile Val Thr
 100 105 110
 Val His Pro Cys Pro Asp Ile Lys Tyr Ala Asn Arg Ile Ser Val Leu
 115 120 125
 Pro Ile Ala Asp Thr Val Glu Gly Ile Asn Gly Ser Leu Phe Asp Leu
 130 135 140
 Tyr Leu Lys Pro Tyr Phe Val Glu Ala Tyr Arg Pro Val Arg Lys Gly
 145 150 155 160
 Asp Leu Phe Thr Val Arg Gly Gly Met Arg Gln Val Glu Phe Lys Val
 165 170 175
 50 Val Glu Val Asp Pro Glu Glu Ile Ala Ile Val Ala Gln Asp Thr Ile
 180 185 190

EP 0 982 401 A2

1 Ile His Cys Glu Gly Glu Pro Ile Asn Arg Glu Asp Glu Glu Asn Ser
 195 200 205
 2 Leu Asn Glu Val Gly Tyr Asp Asp Ile Gly Gly Cys Lys Lys Gln Met
 210 215 220
 5 3 Ala Gln Ile Arg Glu Leu Val Glu Leu Pro Leu Arg His Pro Gln Leu
 225 230 235 240
 4 Phe Lys Ser Ile Gly Ile Lys Pro Pro Lys Gly Ile Leu Met Tyr Gly
 245 250 255
 10 5 Pro Pro Gly Thr Gly Lys Thr Ile Met Ala Arg Ala Val Ala Asn Glu
 260 265 270
 6 Thr Gly Ala Phe Phe Leu Ile Asn Gly Pro Glu Ile Met Ser Lys
 275 280 285
 15 7 Met Ala Gly Glu Ser Glu Ser Asn Leu Arg Lys Ala Phe Glu Glu Ala
 290 295 300
 8 Glu Lys Asn Ser Pro Ser Ile Ile Phe Ile Asp Glu Ile Asp Ser Ile
 305 310 315 320
 20 9 Ala Pro Lys Arg Asp Lys Thr Asn Gly Glu Val Glu Arg Arg Val Val
 325 330 335
 10 Ser Gln Leu Leu Thr Leu Met Asp Gly Met Lys Ala Arg Ser Asn Val
 340 345 350
 11 Val Val Ile Ala Ala Thr Asn Arg Pro Asn Ser Ile Asp Pro Ala Leu
 355 360 365
 25 12 Arg Arg Phe Gly Arg Phe Asp Arg Glu Val Asp Ile Gly Val Pro Asp
 370 375 380
 13 Ala Glu Gly Arg Leu Glu Ile Leu Arg Ile His Thr Lys Asn Met Lys
 385 390 395 400
 14 14 Leu Ala Asp Asp Val Asp Leu Glu Ala Ile Ala Ser Glu Thr His Gly
 405 410 415
 15 Phe Val Gly Ala Asp Ile Ala Ser Leu Cys Ser Glu Ala Ala Met Gln
 420 425 430
 16 16 Gln Ile Arg Glu Lys Met Asp Leu Ile Asp Leu Glu Glu Glu Thr Ile
 435 440 445
 17 Asp Thr Glu Val Leu Asn Ser Leu Gly Val Thr Gln Asp Asn Phe Arg
 450 455 460
 18 Phe Ala Leu Gly Asn Ser Asn Pro Ser Ala Leu Arg Glu Thr Val Val
 465 470 475 480
 40 19 Glu Asn Val Asn Val Thr Trp Asp Asp Ile Gly Gly Leu Asp Asn Ile
 485 490 495
 20 Lys Asn Glu Leu Lys Glu Thr Val Glu Tyr Pro Val Leu His Pro Asp
 500 505 510
 21 21 Gln Tyr Gln Lys Phe Gly Leu Ala Pro Thr Lys Gly Val Leu Phe Phe
 515 520 525
 22 Gly Pro Pro Gly Thr Gly Lys Thr Leu Leu Ala Lys Ala Val Ala Thr
 530 535 540
 23 23 Glu Val Ser Ala Asn Phe Ile Ser Val Lys Gly Pro Glu Leu Leu Ser
 545 550 555 560
 24 Met Trp Tyr Gly Glu Ser Glu Ser Asn Ile Arg Asp Ile Phe Asp Lys
 565 570 575

EP 0 982 401 A2

Ala Arg Ala Ala Ala Pro Thr Val Val Phe Leu Asp Glu Leu Asp Ser
 580 585 590
 Ile Ala Lys Ala Arg Gly Gly Ser His Gly Asp Ala Gly Gly Ala Ser
 595 600 605
 5 Asp Arg Val Val Asn Gln Leu Leu Thr Glu Met Asp Gly Met Asn Ala
 610 615 620
 Lys Lys Asn Val Phe Val Ile Gly Ala Thr Asn Arg Pro Asp Gln Ile
 625 630 635 640
 10 Asp Pro Ala Leu Leu Arg Pro Gly Arg Leu Asp Gln Leu Ile Tyr Val
 645 650 655
 Pro Leu Pro Asp Glu Pro Ala Arg Leu Ser Ile Leu Gln Ala Gln Leu
 660 665 670
 15 Arg Asn Thr Pro Leu Glu Pro Gly Leu Asp Leu Asn Glu Ile Ala Lys
 675 680 685
 Ile Thr His Gly Phe Ser Gly Ala Asp Leu Ser Tyr Ile Val Gln Arg
 690 695 700
 Ser Ala Lys Phe Ala Ile Lys Asp Ser Ile Glu Ala Gln Val Lys Ile
 705 710 715 720
 20 Asn Lys Ile Lys Glu Glu Lys Glu Lys Val Lys Thr Glu Asp Val Asp
 725 730 735
 Met Lys Val Asp Glu Val Glu Glu Asp Pro Val Pro Tyr Ile Thr
 740 745 750
 25 Arg Ala His Phe Glu Glu Ala Met Lys Thr Ala Lys Arg Ser Val Ser
 755 760 765
 Asp Ala Glu Leu Arg Arg Tyr Glu Ser Tyr Ala Gln Gln Leu Gln Ala
 770 775 780
 30 Ser Arg Gly Gln Phe Ser Ser Phe Arg Phe Asn Glu Asn Ala Gly Ala
 785 790 795 800
 Thr Asp Asn Gly Ser Ala Ala Gly Ala Asn Ser Gly Ala Ala Phe Gly
 805 810 815
 Asn Val Glu Glu Asp Asp Leu Tyr Ser
 820 825
 35 (2) INFORMATION FOR SEQ ID NO: 47:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1918 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 40 (ii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO
 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:
 TTTTTTTTTC TCCCTCTCTC TCGTTCAAGAT TCTGTAGAAT TGATTGGTTG AGAGTAAAG 60
 TCAGACTTTT TTTTTTGCTC TCCATCTAGT GGGACAAATA AGAAGTTAA CAAAGAACGA 120
 50 CAAAAAAATCC TCACCAGAAG AAAAAAAAT CAATTTCAC AGGTAAAGTT GTACGGACAG 180
 CACGACAGAC ACAAAACTAA AGTAAATCCA TGAGGAAAAA AGTAAAAAAA AAAAAATTGT 240

EP 0 982 401 A2

5	TCACCACAAAC TTCAAGAGCC ATTAACCA AAAATTTGGA ATATAAATT CAACTGATTT	300
	CTTGCTGGAT TTTTTGTAT ATATTCGAA TTGATTCCT TTTACTTTT TTTTTCCAT	360
	TTCTTCTTT CCTTTTCCA TCTTTAAGT TTCTTTAGA ATATAGTATA TTATCAAAC	420
	AATGTCTGCA TTCAAGATCAA TTCAACGTTA AACCAACGTA GCCAAGAGCA CTTCAAAAA	480
	CAGCATCAGA ACATATGCTT CTGCTGAACC AGTATGTATT CACTTTTG AGGATCCGG	540
10	CAATGTGCTT GGGATTTAC TTTAACGTA TATACAAAGA TAATTTACTA ACTTGCTTC	600
	TTAGACCTTA AAACAAAGAT TGGAAAGAAT CTTGCCAGCC AAAGCTGAAG AAGTTAAACA	660
	ATTCAAAAAA GAACACGGTA AAACGTCTAT TGGTGAAGTT TTATTAAC AAGCTTACGG	720
	TGGTATGAGA GGTATCAAAG GTTGTAGTTG GGAAGGTTCT GTTTGGACCA CAATTGAAGG	780
15	TATCCGTTTC AGAGGAAGAA CCATCCCAGA CATTCAAAAAA GAATTGCCAA AAGCACCAGG	840
	TGGTGAAGAA CCATTACCAAG AAGCTCTTT CTGGTTGTTG TTGACTGGTG AAGTTCCAAC	900
	TGACGCCAA ACTAAGGCTT TATCCGAAGA ATTTGCTGCT AGATCAGGAT TACCAAAAGCA	960
20	CGTTGAAGAA TTGATCGACA GATCTCCATC TCACTTGAC CCAATGGCTC AATTCTCCAT	1020
	TGCCGTTACT GCTTTGGAAT CTGAATCCCA ATTTGCCAA GCTTATGCTA AAGGTGCCAA	1080
	CAAATCCGAA TACTGGAAAT ACACCTACGA AGATTCCATC GATTGTTAG CTAAATTGCC	1140
25	AACCATTGCT GCTAAGATT ACAGAAACGT TTTCCACGAT GGTAAATGTC CAGCTGCCAT	1200
	TGACTCCAAA TTGGATTACG GTGCTAACCTT GGCCAGTTG TTAGGTTTG GTGACAACAA	1260
	GGAATTGTT GAATTAATGA GATTGTACCT TACCATCCAC TCTGACCAAGG AAGGTGGTAA	1320
30	CGTCTCTGCA CACACCACCC ACTTGGTTGG TTCCGCTTTA TCTTCCCCAT TCTTGTCTT	1380
	AGCTGCTGGT TTGAATGGTT TAGCTGGTCC ATTACACGGT AGAGCTAAC AAGAAGTTT	1440
	GGAATGGTTG TTCAAATTAA GAGAAGAATT AAACGGTGAC TACTCCAAGG AAGCCATTGA	1500
35	AAAATACTTG TGGGAAACCT TGAACCTCCGG TAGAGTTGTC CCAGGTTACG GTCACGCTGT	1560
	CTTGAGAAAG ACCGATCCAA GATAACTGTC TCAAAAGAGAA TTTGCTCTTA AACATATGCC	1620
	AGACTACGAA TTGTTCAAAT TGGTTCAAAT CATTACGAA GTCGCTCCAG GTGTTTTAAC	1680
	CAAACACGGT AAGACCAAGA ACCCATGGCC AAATGTGGAC TCCCACCTCG GTGTCTTGT	1740
40	ACAATACTAC GGTTGACTG AACAACTTT CTACACTGTC TTGTTGGTG TTTCCAGAGC	1800
	CTTTGGTGTGTC TTGCCACAAT TGATCTTGGA CCGTGGTATC GGTATGCCAA TTGAAAGACC	1860
	AAAATCTTTC TCCACTGAAA AATACATTGA ATTGGTCAA AACATCAACA AAGCTTAA	1918

45 (2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 466 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

50 (ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

	Met Ser Ala Phe Arg Ser Ile Gln Arg Ser Thr Asn Val Ala Lys Ser
1	5 10 15
5	Thr Phe Lys Asn Ser Ile Arg Thr Tyr Ala Ser Ala Glu Pro Thr Leu
	20 25 30
	Lys Gln Arg Leu Glu Glu Ile Leu Pro Ala Lys Ala Glu Glu Val Lys
	35 40 45
10	Gln Phe Lys Lys Glu His Gly Lys Thr Val Ile Gly Glu Val Leu Leu
	50 55 60
	Glu Gln Ala Tyr Gly Gly Met Arg Gly Ile Lys Gly Leu Val Trp Glu
	65 70 75 80
	Gly Ser Val Leu Asp Pro Ile Glu Gly Ile Arg Phe Arg Gly Arg Thr
	85 90 95
15	Ile Pro Asp Ile Gln Lys Glu Leu Pro Lys Ala Pro Gly Gly Glu Glu
	100 105 110
	Pro Leu Pro Glu Ala Leu Phe Trp Leu Leu Leu Thr Gly Glu Val Pro
	115 120 125
20	Thr Asp Ala Gln Thr Lys Ala Leu Ser Glu Glu Phe Ala Ala Arg Ser
	130 135 140
	Ala Leu Pro Lys His Val Glu Glu Leu Ile Asp Arg Ser Pro Ser His
	145 150 155 160
	Leu His Pro Met Ala Gln Phe Ser Ile Ala Val Thr Ala Leu Glu Ser
	165 170 175
25	Glu Ser Gln Phe Ala Gln Ala Tyr Ala Lys Gly Ala Asn Lys Ser Glu
	180 185 190
	Tyr Trp Lys Tyr Thr Tyr Glu Asp Ser Ile Asp Leu Leu Ala Lys Leu
	195 200 205
30	Pro Thr Ile Ala Ala Lys Ile Tyr Arg Asn Val Phe His Asp Gly Lys
	210 215 220
	Leu Pro Ala Ala Ile Asp Ser Lys Leu Asp Tyr Gly Ala Asn Leu Ala
	225 230 235 240
	Ser Leu Leu Gly Phe Gly Asp Asn Lys Glu Phe Val Glu Leu Met Arg
	245 250 255
35	Leu Tyr Leu Thr Ile His Ser Asp His Glu Gly Gly Asn Val Ser Ala
	260 265 270
	His Thr Thr His Leu Val Gly Ser Ala Leu Ser Ser Pro Phe Leu Ser
	275 280 285
40	Leu Ala Ala Gly Leu Asn Gly Leu Ala Gly Pro Leu His Gly Arg Ala
	290 295 300
	Asn Gln Glu Val Leu Glu Trp Leu Phe Lys Leu Arg Glu Glu Leu Asn
	305 310 315 320
45	Gly Asp Tyr Ser Lys Glu Ala Ile Glu Lys Tyr Leu Trp Glu Thr Leu
	325 330 335
	Asn Ser Gly Arg Val Val Pro Gly Tyr Gly His Ala Val Leu Arg Lys
	340 345 350
	Thr Asp Pro Arg Tyr Thr Ala Gln Arg Glu Phe Ala Leu Lys His Met
50	355 360 365
	Pro Asp Tyr Glu Leu Phe Lys Leu Val Ser Asn Ile Tyr Glu Val Ala

EP 0 982 401 A2

	370	375	380													
	Pro	Gly	Val	Leu	Thr	Lys	His	Gly	Lys	Thr	Lys	Asn	Pro	Trp	Pro	Asn
5	385					390				395				400		
	Val	Asp	Ser	His	Ser	Gly	Val	Leu	Leu	Gln	Tyr	Tyr	Gly	Leu	Thr	Glu
						405				410				415		
	Gln	Ser	Phe	Tyr	Thr	Val	Leu	Phe	Gly	Val	Ser	Arg	Ala	Phe	Gly	Val
10						420				425				430		
	Leu	Pro	Gln	Leu	Ile	Leu	Asp	Arg	Gly	Ile	Gly	Met	Pro	Ile	Glu	Arg
						435				440				445		
	Pro	Lys	Ser	Phe	Ser	Thr	Glu	Lys	Tyr	Ile	Glu	Leu	Val	Lys	Asn	Ile
20		450				455				460						
	Asn	Lys														
15		465														

(2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 678 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:						
	TTTCTGATT	ATCATGTTAT	TTGGTTAGCT	AAACGGAATA	ATGGGATAAT	GGAAGCTGAA	60
	TATCGATTAT	ATTATTAGT	TATCACTTA	ATCATTTCAC	CCGTAGGGTT	AATTATGTTT	120
30	GGTGGTGGTG	CCGCTAGAGA	ATGGCCATGG	CAAGTGATT	ATGTTGGATT	AGGTTTCATT	180
	GGGTTGGTT	GGGGATCAAT	TGGTGATACT	TCAATGTCTT	ATTTAATGGA	TGCTTATCCT	240
	GATATTGTCA	TTCAAGGAAT	GGTGGGAGTA	AGTATTATTA	ATAATAC	GGCTTGATT	300
35	TTCAC	TTTGTTCTTA	TTGGTTACAT	GGATCAGGAA	CACAAAACAC	ATATATTGCC	360
	TTGTC	TTGATTTGC	TACCATAGCA	TTGGTTTTCC	CCTTTTATA	TTATGGTAAA	420
	ACATT	TAGAA	GGAAA	ACTAA	AAGACTTTAT	GTTCAATGG	480
	GGATA	AGAGA	GTGAGTGGTA	AAA	AGAATTTT	ATTAATGATA	540
40	ACTATGGAAA	TCCGAGTCTG	TGTTTTTTT	AGAAGTATAT	TTTAGACGTA	TTTAGAGTTG	600
	TTTTCTCCT	TTGTACTTTA	TTAGCATT	TATAATATAT	TAATTCAAGT	TGCATTAATA	660
	TATATAAATA	AAAAAACT					678

(2) INFORMATION FOR SEQ ID NO: 50:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 159 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Ser Asp Tyr His Val Ile Trp Leu Ala Lys Arg Asn Asn Gly Ile Met
1 5 10 15

5 Glu Ala Glu Tyr Arg Leu Tyr Leu Leu Val Ile Thr Leu Ile Ile Ser
20 25 30

Pro Val Gly Leu Ile Met Phe Gly Val Gly Ala Ala Arg Glu Trp Pro
35 40 45

10 Trp Gln Val Ile Tyr Val Gly Leu Gly Phe Ile Gly Phe Gly Trp Gly
50 55 60

Ser Ile Gly Asp Thr Ser Met Ser Tyr Leu Met Asp Ala Tyr Pro Asp
65 70 75 80

15 Ile Val Ile Gln Gly Met Val Gly Val Ser Ile Ile Asn Asn Thr Leu
85 90 95

Ala Cys Ile Phe Thr Phe Ala Cys Ser Tyr Trp Leu Asp Gly Ser Gly
100 105 110

20 Thr Gln Asn Thr Tyr Ile Ala Leu Ser Ile Ile Asp Phe Ala Thr Ile
115 120 125

Ala Leu Val Phe Pro Phe Leu Tyr Tyr Gly Lys Thr Phe Arg Arg Lys
130 135 140

Thr Lys Arg Leu Tyr Val Ser Met Val Glu Leu Thr Gln Gly Met
145 150 155

25 (2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1480 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

35 (ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1060
- (D) OTHER INFORMATION:/note= "R = A or G"

40 (ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1063
- (D) OTHER INFORMATION:/note= "Y = C or T"

45 (ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1123
- (D) OTHER INFORMATION:/note= "Y = C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

50 TTTGGATTTT CAATTACAAG ATATTTTGCA TCATGTTGAA AGCAAATGGT TTGGTGGGTT 60

TATTTCAAGGT ATTTCACTA ATGACAATGA CGTTGAAAAT GAATCCAAGA ACGTGTTCA 120

TAAATTCAAA CAAGATTAA TGAAAATTTT GAAAGATTGT TTAACCGTAA GTGACGATAA 180

EP 0 982 401 A2

	ATCGAATATA GAGAGGTTTC TTCAGTTAA TGAATTATT TATTACTGCT TTTACTCAAT	240
	GGAGGAATAT AATTATGAAT TGGTTGATGA TTTGATAAAA TTTATAACTA TAAATATGAA	300
5	TTCTCATGGC AGAATAGTTA ATTTGGCAC TAATGTTAAA ATTAATAAT TACACGAATT	360
	AATTAAGAAT TTGATTGATA AAGTTAATAA AAACAAACAA AATGTGACTA GCAACAACAA	420
	AAACAAACAAAC AACAACAACA GCAACAACAA CAGCAACAGC AACAAATTCCC AACATATTGT	480
	TTTGATACTT AATGCCAACT GTTCCAATT CCCATGGGAA TCGATGGAAT TTCTTCGTAG	540
10	TAAATCAATT TCAAGAATGC CATCAATTCA TATGTTACTT GATCTAGTCA AATCAAACAC	600
	CAATAACAG AACAAGTTAA TGTTTGTGA TAAATCTAAT TTGTATTATT TGATTAATCC	660
	CAGTGGTGTAT TAAATTCGAT CAGAAAATCG ATTCAAAAAA CTATTTGAAT CAAATCATT	720
15	ATGGAGAGGG GAAATTGGAA AATTATCAAG TAATGAACAT GAAGATTATC AAGATTCAAT	780
	ATTATGTGAA ATCTTGAAAAA GTCAATTATT TGTTTATATT GGTCTATGGT GTTGTGATCA	840
	ATATATTAAA GTATCAAAAT TATTTAAAAA ATGTGGCAAT AATCAAGATT TACTGAATAA	900
	ATTACCTCCT AGTTTATTGT TAGGTTGTTC ATCAGTTAAA TTAGATAATT GTATTATAA	960
20	CTATAATTCC AGTATGTTAC ACCACTGGG TAATATTAT AATTGGTTGA ACTGTAATC	1020
	GTCAATGATA CTCGGGAATC TATGGGATGT TACTGATAAR GAYATTGATA TTTTACACT	1080
	TTCATTACTA CAAAATGGG GGTTAATAGA TGATTATAAT GGYAGTGGCC ATGATTATGG	1140
25	TATGAAGAAA TTGGATTGTA CTAATTGTGT TGTTCAAAAGT CGAAGTAAAT GTACTTTGAA	1200
	ATACATTGAAT GGATCAGCAC CTGTGGTTA TGGTCTACCA ATGTATTAA AATAGACATT	1260
	CTGTTGCAAT ATAAGTTTAT ATATTTAAAT AATAAGAAAA AGAGCATAAT TTGGATCTTG	1320
	ATTTTGTATT GTTGGTTTG TTATGAACAA ATTTGCACC CAATCACTAT CGAACCTTCT	1380
30	TTTTAAACA GAGAACATT AATCAACATT TATGTTACAT TTAAGCGTTT AAATACATAT	1440
	TTGTGTTAGA TACTTATATA ATGTTGATG CAAACATACA	1480

(2) INFORMATION FOR SEQ ID NO: 52:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 417 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

40 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

45	Leu Asp Phe Gln Leu Gln Asp Ile Leu His His Val Glu Ser Lys Trp	
	1 5 10	15
	Phe Gly Gly Phe Ile Ser Gly Ile Phe Thr Asn Asp Asn Asp Val Glu	
	20 25 30	
50	Asn Glu Ser Lys Asn Val Phe His Lys Phe Lys Gln Asp Leu Met Lys	
	35 40 45	
	Ile Leu Lys Asp Cys Leu Thr Val Ser Asp Asp Lys Ser Asn Ile Glu	

EP 0 982 401 A2

	50	55	60
	Arg Phe Leu Gln Phe Asn Glu Phe Ile Tyr Tyr Cys Phe Tyr Ser Met		
65	65 70	75	80
5	Glu Glu Tyr Asn Tyr Glu Leu Val Asp Asp Leu Ile Lys Phe Ile Thr		
	85	90	95
	Ile Asn Met Asn Ser His Gly Arg Ile Val Asn Phe Gly Thr Asn Val		
	100	105	110
10	Lys Ile Asn Lys Leu His Glu Leu Ile Lys Asn Leu Ile Asp Lys Val		
	115	120	125
	Asn Lys Asn Lys Gln Asn Val Thr Ser Asn Asn Lys Asn Asn Asn Asn		
	130	135	140
	Asn Asn Ser Asn Asn Ser Asn Ser Asn Asn Ser Gln His Ile Val		
15	145 150	155	160
	Leu Ile Pro Asn Ala Asn Cys Ser Asn Phe Pro Trp Glu Ser Met Glu		
	165	170	175
	Phe Leu Arg Ser Lys Ser Ile Ser Arg Met Pro Ser Ile His Met Leu		
	180	185	190
20	Leu Asp Leu Val Lys Ser Asn Thr Asn Asn Lys Asn Lys Leu Met Phe		
	195	200	205
	Val Asp Lys Ser Asn Leu Tyr Tyr Leu Ile Asn Pro Ser Gly Asp Leu		
	210	215	220
25	Ile Arg Ser Glu Asn Arg Phe Lys Lys Leu Phe Glu Ser Asn His Leu		
	225 230	235	240
	Trp Arg Gly Glu Ile Gly Lys Leu Ser Ser Asn Glu His Glu Asp Tyr		
	245 250	255	
	Gln Asp Ser Ile Leu Cys Glu Ile Leu Lys Ser His Leu Phe Val Tyr		
	260	265	270
30	Ile Gly His Gly Gly Cys Asp Gln Tyr Ile Lys Val Ser Lys Leu Phe		
	275 280	285	
	Lys Lys Cys Gly Asn Asn Gln Asp Leu Ser Asn Lys Leu Pro Pro Ser		
	290	295	300
35	Leu Leu Leu Gly Cys Ser Ser Val Lys Leu Asp Asn Cys Asn Tyr Asn		
	305 310	315	320
	Tyr Asn Ser Ser Met Leu Gln Pro Ser Gly Asn Ile Tyr Asn Trp Leu		
	325 330	335	
40	Asn Cys Lys Ser Ser Met Ile Leu Gly Asn Leu Trp Asp Val Thr Asp		
	340 345	350	
	Xaa Xaa Ile Asp Ile Phe Thr Leu Ser Leu Leu Gln Lys Trp Gly Leu		
	355 360	365	
	Ile Asp Asp Tyr Asn Xaa Ser Gly His Asp Tyr Gly Met Lys Lys Leu		
	370 375	380	
45	Asp Leu Thr Asn Cys Val Val Gln Ser Arg Ser Lys Cys Thr Leu Lys		
	385 390	395	400
	Tyr Leu Asn Gly Ser Ala Pro Val Val Tyr Gly Leu Pro Met Tyr Leu		
	405 410	415	
50	Lys		

(2) INFORMATION FOR SEQ ID NO: 53:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1443 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

CTTCTTTAG AGACAATGCA GTGGTTTCT TACCAGATGC ATGACCCCA CCCAATAAAA	60
CTATAATCGA TCTATTCA CA GTATTTGATG CCATTTGAT GGTGATGAAT GATGTGATGT	120
GATGCTCATC TTATTGGGAG TTCAAAAAAA AAAAGTTACA CTCGAAAAAA AAAAATAGC	180
ATTATAAATA GAAGCTTAC TATCTTATAG AACAAAACAA AAAACACTAT CTTCTAATTA	240
ATAATGGATG ATTTGATAG AGATTTAGAT AATGAGTTGG AATTTAGTCA TAAATCAACG	300
AAAGGAATAA AGGTTCATCG CACTTTGAA ACTATGAATT TGAAACCTGA TCTTTGAAA	360
GGAATATATG CCTATGGATT TGAACCACCA TCTGCTATTG AATCTAGGGC TATTATGCAG	420
ATCATCAGTG GTAGAGACAC AATAGCACAG GCACAATCTG GAACTGGTAA AACTGCTACT	480
TTTTCTATTG GTATGCTTGA GTTATAGAT ACTAAATCAA AAGAGTGTCA AGCACTTATC	540
TTGTCCTCA TAAGAGAGTT GGCAATTCAA ATACAAAATG TGGTCATGCA TTTAGGAGAT	600
TATATGAACA TTCACACCCA TGCCTGTATT GGTGGAAAAA ATGTCGGTGA GGATGTTAAG	660
AAATTGCAGC AAGGGCAACA AATAGTTAGT GGGACACCG AGTAGAGTGT TGATGTGATA	720
AAAAGAAGAA ATCTACAAAC TAGAAATATC AAGGTTCTTA TTTAGATGA AGCTGATGAA	780
CTTTTACAA AAGGGTTAA AGAACAGATC TACGAAATCT ACAAACACATT ACCACCTTCG	840
GTTCAAGTAG TAGTTGTTAG TGCCACTTTG CCACGTGAAG TATTGGAGAT GACAAGTAAG	900
TTTACCACTG ATCCAGTGAA AATCTTGGTG AAGAGGGATG AGATTCGCT TCTGGGAATC	960
AAACAAATATT ATGTTCAATG TGAACGTGAA GATTGGAAGT TTGATACACT ATGTGATTG	1020
TATGACAACC TTACAATAAC TCAAGCAGTG ATATTTGTA ATACCAAATT GAAGGTGAAT	1080
TGGCTTGCTG ATCAAATGAA AAAGCAAAAC TTTACTGTTG TGGCAATGCA TGGTGATATG	1140
AAACAAGATG AACGAGATTC AATTATGAAC GATTTAGAA GGGGGAAATTC AAGAGTATTA	1200
ATATCTACAG ATGTTGGGC AAGAGGTATT GATGTCCAAAC AAGTCTCGTT GGTAATAAT	1260
TATGATTG CCAACCGATAA GGAAAATCT ATTCAATAGAA TTGGACGATC AGGTAGATTT	1320
GGTAGAAAGG GAAACAGCTAT AAACCTGATA ACTAAAGATG ATGTGGTCAC TTTAAAGAA	1380
TTGGAGAAAT ATTATCAAC GAAAATTAAG GARATGCCAA TGAATATTAA TGATATAATG	1440
TAA	1443

45 (2) INFORMATION FOR SEQ ID NO: 54:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 399 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:

EP 0 982 401 A2

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

10 Met Asp Asp Phe Asp Arg Asp Leu Asp Asn Glu Leu Glu Phe Ser His
1 5 10 15
Lys Ser Thr Lys Gly Ile Lys Val His Arg Thr Phe Glu Ser Met Asn
20 25 30
Leu Lys Pro Asp Leu Leu Lys Gly Ile Tyr Ala Tyr Gly Phe Glu Ala
35 40 45
15 Pro Ser Ala Ile Gln Ser Arg Ala Ile Met Gln Ile Ile Ser Gly Arg
50 55 60
Asp Thr Ile Ala Gln Ala Gln Ser Gly Thr Gly Lys Thr Ala Thr Phe
65 70 75 80
20 Ser Ile Gly Met Leu Glu Val Ile Asp Thr Lys Ser Lys Glu Cys Gln
85 90 95
Ala Leu Ile Leu Ser Pro Thr Arg Glu Leu Ala Ile Gln Ile Gln Asn
100 105 110
Val Val Met His Leu Gly Asp Tyr Met Asn Ile His Thr His Ala Cys
115 120 125
25 Ile Gly Gly Lys Asn Val Gly Glu Asp Val Lys Lys Leu Gln Gln Gly
130 135 140
Gln Gln Ile Val Ser Gly Thr Pro Gly Arg Val Ile Asp Val Ile Lys
145 150 155 160
30 Arg Arg Asn Leu Gln Thr Arg Asn Ile Lys Val Leu Ile Leu Asp Glu
165 170 175
Ala Asp Glu Leu Phe Thr Lys Gly Phe Lys Glu Gln Ile Tyr Glu Ile
180 185 190
Tyr Lys His Leu Pro Pro Ser Val Gln Val Val Val Ser Ala Thr
195 200 205
35 Leu Pro Arg Glu Val Leu Glu Met Thr Ser Lys Phe Thr Thr Asp Pro
210 215 220
Val Lys Ile Leu Val Lys Arg Asp Glu Ile Ser Leu Ser Gly Ile Lys
225 230 235 240
40 Gln Tyr Tyr Val Gln Cys Glu Arg Glu Asp Trp Lys Phe Asp Thr Leu
245 250 255
Cys Asp Leu Tyr Asp Asn Leu Thr Ile Thr Gln Ala Val Ile Phe Cys
260 265 270
45 Asn Thr Lys Leu Lys Val Asn Trp Leu Ala Asp Gln Met Lys Lys Gln
275 280 285
Asn Phe Thr Val Val Ala Met His Gly Asp Met Lys Gln Asp Glu Arg
290 295 300
50 Asp Ser Ile Met Asn Asp Phe Arg Arg Gly Asn Ser Arg Val Leu Ile
305 310 315 320
Ser Thr Asp Val Trp Ala Arg Gly Ile Asp Val Gln Gln Val Ser Leu

55

EP 0 982 401 A2

	325	330	335
	Val Ile Asn Tyr Asp Leu Pro Thr Asp Lys Glu Asn Tyr Ile His Arg		
	340	345	350
5	Ile Gly Arg Ser Gly Arg Phe Gly Arg Lys Gly Thr Ala Ile Asn Leu		
	355	360	365
	Ile Thr Lys Asp Asp Val Val Thr Leu Lys Glu Leu Glu Lys Tyr Tyr		
	370	375	380
10	Ser Thr Lys Ile Lys Glu Met Pro Met Asn Ile Asn Asp Ile Met		
	385	390	395

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1020 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

	AACGTTGCC TGGCCCAAGTT ATTCCGTTT CCAAGCAAAAT GAATGTCGAT ACCGACATCA	60
25	TCACGTTGAC CCGTTTTATT TTACAAGAAC AGCAAACGTG TGCTCCACC GCCACCGGTG	120
	AGTTGTCGTT GTTGTGAAT GCGCTTCAT TTGCATTCAA GTTTATTGCC CACAATATCA	180
	GAAGAGCTGA GTTGGTCAAC CTTATTGGTG TTTCTGGCTC TGCCAACCTCT ACCGGTGATG	240
30	TTCAGAAGAA ATTGGATGTG ATTGGTGATG AGATCTTAT CAATGCCATG AGATCTTCCA	300
	ACAACGTCAA GGTGTTGGTT TCTGAAGAGC AAGAAGACCT TATTGTGTC CCAGGTGGTG	360
	GCACATATGC TGTTGTACT GATCCAATTG ATGGGTCGTC CAATATCGAT GCTGGTGT	420
35	CTGTTGGTAC GATTTGGTGT GTGTACAAGT TGCAAGAGGG GTCTACTGGT GGCATCAGCG	480
	ATGTCCTGGC TCTCGTAAG GAGATGGTCG CTGGGGGTA CACCATGTAC GGTGCATCTG	540
	CCCATTTGGC ATTGACTACA GGTCACTGGT TCAATCTTT TACTTTGGAT ACTCAGTTGG	600
	GTGAATTAT CTTGACCCAT CCAAACATTGA AGTTGCCAGA TACTAAGAAC ATCTACTCGT	660
40	TGAATGAAGG GTACTCGAAC AAATTCCAG AATACGTTCA AGATTATCTG AAGGACATTA	720
	AAAAGGAAGG GTACAGTTG AGATACATTG GACTGATGGT TGCTGATGTC CATCGTACTC	780
	TTTTGTATGG TGGTATTTT GCTTACCTA CATTAAAGTT GAGAGTGTG TATGAATGTT	840
	TCCCCATGGC CTTGTTGATG GAACAAGCAG GCGGTTCTGC TGTCACCATC AAGGGTGAGA	900
45	GGATCTTGGGA TATCTTGCCA AAAGGTATAC ACGACAAGAG TTCTATTGTG TTGGGATCCA	960
	AGGGTGAAGT TGAAGTAT TAAAGCATG TACCAAAATA GATTATGTAG AAAATTATG	1020

(2) INFORMATION FOR SEQ ID NO: 56:

50

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 320 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:

55

EP 0 982 401 A2

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

10 Met Asn Val Asp Thr Asp Ile Ile Thr Leu Thr Arg Phe Ile Leu Gln
1 5 10 15

Glu Gln Gln Thr Val Ala Pro Thr Ala Thr Gly Glu Leu Ser Leu Leu
20 25 30

Leu Asn Ala Leu Gln Phe Ala Phe Lys Phe Ile Ala His Asn Ile Arg
35 40 45

15 Arg Ala Glu Leu Val Asn Leu Ile Gly Val Ser Gly Ser Ala Asn Ser
50 55 60

Thr Gly Asp Val Gln Lys Lys Leu Asp Val Ile Gly Asp Glu Ile Phe
65 70 75 80

20 Ile Asn Ala Met Arg Ser Ser Asn Asn Val Lys Val Leu Val Ser Glu
85 90 95

Glu Gln Glu Asp Leu Ile Val Phe Pro Gly Gly Thr Tyr Ala Val
100 105 110

25 Cys Thr Asp Pro Ile Asp Gly Ser Ser Asn Ile Asp Ala Gly Val Ser
115 120 125

Val Gly Thr Ile Phe Gly Val Tyr Lys Leu Gln Glu Gly Ser Thr Gly
130 135 140

30 Gly Ile Ser Asp Val Leu Arg Pro Gly Lys Glu Met Val Ala Ala Gly
145 150 155 160

Tyr Thr Met Tyr Gly Ala Ser Ala His Leu Ala Leu Thr Thr Gly His
165 170 175

Gly Val Asn Leu Phe Thr Leu Asp Thr Gln Leu Gly Glu Phe Ile Leu
180 185 190

35 Thr His Pro Asn Leu Lys Leu Pro Asp Thr Lys Asn Ile Tyr Ser Leu
195 200 205

Asn Glu Gly Tyr Ser Asn Lys Phe Pro Glu Tyr Val Gln Asp Tyr Ser
210 215 220

40 Lys Asp Ile Lys Lys Glu Gly Tyr Ser Leu Arg Tyr Ile Gly Ser Met
225 230 235 240

Val Ala Asp Val His Arg Thr Leu Leu Tyr Gly Gly Ile Phe Ala Tyr
245 250 255

45 Pro Thr Leu Lys Leu Arg Val Leu Tyr Glu Cys Phe Pro Met Ala Leu
260 265 270

Leu Met Glu Gln Ala Gly Gly Ser Ala Val Thr Ile Lys Gly Glu Arg
275 280 285

Ile Leu Asp Ile Leu Pro Lys Gly Ile His Asp Lys Ser Ser Ile Val
290 295 300

50 Leu Gly Ser Lys Gly Glu Val Glu Lys Tyr Leu Lys His Val Pro Lys
305 310 315 320

55

EP 0 982 401 A2

(2) INFORMATION FOR SEQ ID NO: 57:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LEN TH: 825 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

15 AACCCACCT TCAAGACAA AGAAGATTTC GTCAAGAAA CGAATGTCAG AGCAGAAAAG 60
ACCAAGAAC TAATCAAATT TGCCCGTGAC AACCTAACC ATTTACCAATT CACCGAAAAA 120
GACGGAGGTG CATGGAAAAA CTATGAACGA ATGATCAGTG GTATGCTCA CAACTGTTA 180
CRAAAAAGAAT TGGAACAAAC ACGTATGTCT TGCAAGACT ACATGTTGA CTACGGCAGT 240
TTCAGAACTA GAGATTATAA ACAACCCAA GAATTTCTTG ATGCAAAATA CAAACATTAA 300
20 GAAAGTTCA TTGGACATGT TGGAAAAAT GCATTTATGG AATATCCAAT CTATTTGAT 360
TATGGTTTA ACACCTATTT GGGTGATAAT TTCTATTCCA ATTACAATT GACAATTTG 420
GATGTTCCA TAGTCAGAAT TGGTATAAT GTCAAGTGTG GTCCCAATGT ATCTATCCTT 480
25 ACCCAACAC ACCCAGTGGA TCCCACTTG CGCTATGATC AATTGGAAAA TGCCCTGCCT 540
GTGACGGTGG GTAACGGGGT CTGGTTGTGT GGAAGCTGTA CCATTCTTGG TGGGGTGACA 600
GTAGGTGATG GCAGCATTGT GGCTGCTGGT GCAGTTGTCA ACAAGGACGT TCCACCAAAC 660
30 ACTGTAGTTG CGGGAGTTCC TGCTAGGGTA GTTAAGCAGC TAGAACCTAG AGACCCAAAC 720
TTTGACACTA TGGCAGTTT GAAACAATAT GGTATGGTT ATATAGATTA GTAATTAGAT 780
TTGATGTAAT GTACACGACT ACACATTGT CTGGTGCTG TTTT 825

(2) INFORMATION FOR SEQ ID NO: 58:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 206 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

45 Met Ile Ser Gly Met Leu Tyr Asn Cys Leu Gln Lys Glu Leu Glu Thr 15
1 5 10 15
Thr Arg Met Ser Cys Arg Asp Tyr Met Leu Asp Tyr Gly Ser Phe Arg 30
20 25 30
50 Thr Arg Asp Tyr Lys Thr Thr Gln Glu Phe Leu Asp Ala Lys Tyr Lys 45
35 40 45
His Leu Glu Ser Phe Ile Gly His Val Gly Lys Asn Ala Phe Met Glu

EP 0 982 401 A2

	50	55	60	
	Tyr Pro Ile Tyr Phe Asp Tyr Gly Phe Asn Thr Tyr Leu Gly Asp Asn			
	65	70	75	80
5	Phe Tyr Ser Asn Tyr Asn Leu Thr Ile Leu Asp Val Ser Ile Val Arg			
	85	90	95	
	Ile Gly Asn Asn Val Lys Cys Gly Pro Asn Val Ser Ile Leu Thr Pro			
	100	105	110	
10	Thr His Pro Val Asp Pro Thr Leu Arg Tyr Asp Gln Leu Glu Asn Ala			
	115	120	125	
	Leu Pro Val Thr Val Gly Asn Gly Val Trp Leu Cys Gly Ser Cys Thr			
	130	135	140	
	Ile Leu Gly Gly Val Thr Val Gly Asp Gly Ser Ile Val Ala Ala Gly			
	145	150	155	160
15	Ala Val Val Asn Lys Asp Val Pro Pro Asn Thr Val Val Ala Gly Val			
	165	170	175	
	Pro Ala Arg Val Val Lys Gln Leu Glu Pro Arg Asp Pro Asn Phe Asp			
	180	185	190	
20	Thr Met Ala Val Leu Lys Gln Tyr Gly Met Gly Tyr Ile Asp			
	195	200	205	

(2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1380 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

30

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:						
	AATTACAATC	TGGTTGGTA	CTACCATATC	CCATTAGTGT	TATTGTCATT	GTAGATATTG	60
35	ATATGGTTA	AAGGATTGGT	TTTCATTTT	TGTGTAATGA	ATGAGCCAAA	ATAAAAAAATC	120
	AATTGATGC	GATGCAATGA	AGTTTAATAA	AATTTTTTTT	TTTCTTTATT	TCTTTTAATC	180
	AACCCATCAA	TCATTAATT	GAATCAATAC	CTACCATTA	CATACTTCTA	TATACATATA	240
40	TATATATAAC	AAAATATCAT	GGGAAAGATA	ACAACTAGTG	ATACTAAAC	AAAACAACGT	300
	CATAATCCAT	TATTAAGA	TATTCATCC	CAAGGTGGGA	ATTTAAGAAC	CGTTCCAAGA	360
	TCATCATCAT	CATCATCATC	ACAAAAGAAG	AAATCATCAA	AGAAACAAAG	ACATAACGAT	420
	GAAGACGACG	AAGAAAATGG	TGGCGGTGAA	GGATTTTAG	ATGCTTCAG	TTCAAGAAAG	480
45	ATTTACAAT	TGGCAAAAGA	ACAACAAGAT	GAACCTGAAC	AAGAAGATGA	AATACAAAAT	540
	AAACCTTCAT	TTGCTCAATC	ATTTAAAAT	CAACAAATAG	ATAGTGAAGA	AGAAGAAGAG	600
	GAAGATGAGT	ATTCAGATT	TGAAGAAGAA	GAAGAAGTTG	AAGAGATAGT	ATATGATGAA	660
50	GAAGATGCAG	AAGTTGATCC	CAAAGATGCA	GAATTATTAA	ATAAAATATT	CCAATCCAAC	720
	GGTGAAGCTA	ATAATAATGA	TGATGATAAT	TCATTTCAAC	CAACAATAAA	TTTAGCTGAT	780

55

EP 0 982 401 A2

5	AAAATCTTAG CCAAAATTCA AGAAAAAGAA TCCCAACAAAC ACAACAAACA ACAAAGCTCT	840
	CCAGATAATA GTAATGAAGA TGCGTATTG TTACCACCAA AAGTCATTT AGCTTATGAA	900
	AAAATTGGTC AAATTTTATC AACTTATACT CATGGGAAAT TACCTAAATT ATTTAAAATT	960
	TTACCAAGTT TAAAAAATTG GCAAGATGTA TTATACGTGA CAAATCCAAA TAGTTGGACT	1020
	CCTCATGCCA CATATGAAGC AACTAAATTAA TTTGTGTCGA ATTTATCAAG TAATGAAAGCT	1080
10	ACAGTTTCA TTGAAACTAT CTTGTTGCCA CGATTCCGTG ATTCTATTGA AAATTCCGAT	1140
	GATCATTATCAT TAAATTATCA TATTTATCGA GCATTAAGAA ATACATTATA TAAACCAGGA	1200
	GCTTTTTCA AAGGGTTCTT GTTACCTTTA GTCGATGGTT ATTGTTCTGT ACGTGAAGCC	1260
	ACTATTGCTG CTTCACTGTT AACTAAAGTT TCTGTCCCTG TTTTACATTC ATGTCATTAT	1320
	TGTGGCGTAC TGATGAATAA AAAACGAGAA TCACCTGTAT TTGTCCCTACG GCGAATATAA	1380

15

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 373 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

20

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

30	Met Gly Lys Ile Thr Thr Ser Asp Thr Lys Thr Lys Gln Arg His Asn 1 5 10 15
	Pro Leu Leu Lys Asp Ile Ser Ser Gln Gly Gly Asn Leu Arg Thr Val 20 25 30
	Pro Arg Ser Ser Ser Ser Ser Gln Lys Lys Lys Ser Ser Lys 35 40 45
35	Lys Gln Arg His Asn Asp Glu Asp Asp Glu Glu Asn Gly Gly Glu 50 55 60
	Gly Phe Leu Asp Ala Ser Ser Arg Lys Ile Leu Gln Leu Ala Lys 65 70 75 80
	Glu Gln Gln Asp Glu Leu Glu Gln Glu Asp Glu Ile Gln Asn Lys Pro 85 90 95
40	Ser Phe Ala Gln Ser Phe Lys Asn Gln Ile Asp Ser Glu Glu Glu 100 105 110
	Glu Glu Glu Asp Glu Tyr Ser Asp Phe Glu Glu Glu Val Glu 115 120 125
	Glu Ile Val Tyr Asp Glu Glu Asp Ala Glu Val Asp Pro Lys Asp Ala 130 135 140
45	Glu Leu Phe Asn Lys Tyr Phe Gln Ser Asn Gly Glu Ala Asn Asn Asn 145 150 155 160
	Asp Asp Asp Asn Ser Phe Gln Pro Thr Ile Asn Leu Ala Asp Lys Ile 165 170 175
50	Leu Ala Lys Ile Gln Glu Lys Glu Ser Gln Gln Gln Gln Gln

55

EP 0 982 401 A2

	180	185	190
	Ser Ser Pro Asp Asn Ser Asn Glu Asp Ala Val Leu Leu Pro Pro Lys		
	195 200	205	
5	Val Ile Leu Ala Tyr Glu Lys Ile Gly Gln Ile Leu Ser Thr Tyr Thr		
	210 215	220	
	His Gly Lys Leu Pro Lys Leu Phe Lys Ile Leu Pro Ser Leu Lys Asn		
	225 230	235	240
10	Trp Gln Asp Val Leu Tyr Val Thr Asn Pro Asn Ser Trp Thr Pro His		
	245 250	255	
	Ala Thr Tyr Glu Ala Thr Lys Leu Phe Val Ser Asn Leu Ser Ser Asn		
	260 265	270	
	Glu Ala Thr Val Phe Ile Glu Thr Ile Leu Leu Pro Arg Phe Arg Asp		
	275 280	285	
15	Ser Ile Glu Asn Ser Asp Asp His Ser Leu Asn Tyr His Ile Tyr Arg		
	290 295	300	
	Ala Leu Lys Lys Ser Leu Tyr Lys Pro Gly Ala Phe Phe Lys Gly Phe		
	305 310	315	320
20	Leu Leu Pro Leu Val Asp Gly Tyr Cys Ser Val Arg Glu Ala Thr Ile		
	325 330	335	
	Ala Ala Ser Val Leu Thr Lys Val Ser Val Pro Val Leu His Ser Cys		
	340 345	350	
25	His Tyr Cys Gly Val Ser Met Asn Lys Lys Arg Glu Ser Pro Val Phe		
	355 360	365	
	Val Leu Arg Arg Ile		
	370		

(2) INFORMATION FOR SEQ ID NO: 61:

30	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 823 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
35	(iii) HYPOTHETICAL: NO	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:	
40	AACCAACAAT GAGTCAAGTC GCTCCAAAGT GGTACCAATC AGAAGACGTT CCAGCTCAA	60
	AACAAACCAG AAAGACTGCT CGTCCACAAA ATTACGTGC CTCTTTAGTC CCAGGTACCG	120
	TTTTAATTTT ATTGGCCGGT AGATTCAAGAG GTAAAAGAGT TGTTTACTTG AAGAACTTGG	180
45	AAGACAACAC CTTATTGGTT TCTGGTCCAT TCAAAGTCAA TGGTGTCCA TTGAGAAGAG	240
	TTAACGCTAG ATACGTTATC GCCACCTCCA CCAAAGTCAA CGTTTCTGGT GTTGATGTTT	300
	CTAAATTCAA CGTCAATAC TTTGCTAGAG AAAATCTTC TAAATCTAAA AAATCCGAAG	360
	CTGAATTCTT CAATGAATCT CAACCAAAGA AAGAAATCRA AGCTGAAAGA GTTGCTGACC	420
50	AAAAATCTGT CGATGCTGCT TTATTAAGTG AAATCAAAA GACCCCATTA TTGAAACAAT	480
	ACTTGGCCGC TTCATTCTCT TTGAAGAACG GTGACAGACC ACACTTGTTA AAATTTAAT	540

EP 0 982 401 A2

5
TTAGGTGAAA TTAATATTTT GCAACATGT TCATGATAAA TAACAATGTG GCTTTAAAG 600
CAATGGATGG GATATGGTTA AGAGGATGTC TTTATATTTT GAGTTTATA TATGGGTACT 660
TTGTTTAATA ATGGAAGCTA TTGGCTCAGA TGAACCTCAA AATGGAGATT ACTTTTTCT 720
TTTACTTTTA CAATATTTTC GTCTATTGC TGTTTAAGCT GCAAAACAA ATTTTAATC 780
GGTGTATCTT AACTCTTATT CATTGGTAT ATTAAATACA TAT 823

10 (2) INFORMATION FOR SEQ ID NO: 62:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 176 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown

15 (ii) MOLECULE TYPE: peptide

15 (iii) HYPOTHETICAL: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

Met Ser Gln Val Ala Pro Lys Trp Tyr Gln Ser Glu Asp Val Pro Ala
1 5 10 15
Pro Lys Gln Thr Arg Lys Thr Ala Arg Pro Gln Lys Leu Arg Ala Ser
20 25 30
Leu Val Pro Gly Thr Val Leu Ile Leu Ala Gly Arg Phe Arg Gly
35 40 45
Lys Arg Val Val Tyr Leu Lys Asn Leu Glu Asp Asn Thr Leu Val
50 55 60
Ser Gly Pro Phe Lys Val Asn Gly Val Pro Leu Arg Arg Val Asn Ala
65 70 75 80
Arg Tyr Val Ile Ala Thr Ser Thr Lys Val Asn Val Ser Gly Val Asp
85 90 95
Val Ser Lys Phe Asn Val Glu Tyr Phe Ala Arg Glu Lys Ser Ser Lys
100 105 110
Ser Lys Lys Ser Glu Ala Glu Phe Phe Asn Glu Ser Gln Pro Lys Lys
115 120 125
Glu Ile Lys Ala Glu Arg Val Ala Asp Gln Lys Ser Val Asp Ala Ala
130 135 140
Leu Leu Ser Glu Ile Lys Lys Thr Pro Leu Leu Lys Gln Tyr Leu Ala
145 150 155 160
Ala Ser Phe Ser Leu Lys Asn Gly Asp Arg Pro His Leu Leu Lys Phe
165 170 175

45 (2) INFORMATION FOR SEQ ID NO: 63:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

AACATTAAG	CAAGATGGAA	AACGATAAAG	GTCATTAGT	TGAATTATAC	GTCCCAAGAA	60
AATGTTCTGC	TACCAACAGA	ATCATTAAAG	CCAAAGATCA	CGCTTCTGTT	CAAATCTCAA	120
TTGCTAAAGT	TGATGAAGAC	GGTAGAGCTA	TTGCTGGTGA	AAACATCACT	TACGCTTTAA	180
10	GTGGTTACGT	TAGAGGTAGA	GGTGAAGCTG	ATGACTCATT	AAACAGATTG	240
ACGGTTTATT	GAAGAACGTC	TGGTCTTACT	CTCGTTAAGA	GAATAGAAGA	ATAGACAAAA	300
TTGATAATTG	GGTATTTAA	GAATTACTT	TTTTATATT	GCAAATTAAT	TTAAATCTTT	360
15	CTTCTGTGTA	TATTTAATGT	CTTAACATAA	AAAAAAAAAA	GAATAGAAAT	415

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 87 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

Met	Glu	Asn	Asp	Lys	Gly	Gln	Leu	Val	Glu	Leu	Tyr	Val	Pro	Arg	Lys
1															15
Cys	Ser	Ala	Thr	Asn	Arg	Ile	Ile	Lys	Ala	Lys	Asp	His	Ala	Ser	Val
30															30
Gln	Ile	Ser	Ile	Ala	Lys	Val	Asp	Glu	Asp	Gly	Arg	Ala	Ile	Ala	Gly
35															
Glu	Asn	Ile	Thr	Tyr	Ala	Leu	Ser	Gly	Tyr	Val	Arg	Arg	Gly	Glu	
40															
50															
Ala	Asp	Asp	Ser	Leu	Asn	Arg	Leu	Ala	Gln	Gln	Asp	Gly	Leu	Leu	Lys
65															80
Asn	Val	Trp	Ser	Tyr	Ser	Arg									
45															

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1519 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

50

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 749

55

EP 0 982 401 A2

(D) OTHER INFORMATION:/note= "N = A or T or C or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID N : 65:

5	ACCATGTGTC AAATTGCTTG GTCGTGTCCT TTCACCACAC ATTTTTTG ATTAAATTTC	60
	TCGCACGCTC AAAAATGAC TTCGACAAAA AGCAATGCCA CTCTTCCTAC AATTAATTCC	120
	CTCCGCCCCCT TCCTTTCTAT ATACTATCTC CCTTCCTCTC TCCCTCTCCT TTTATTTTT	180
10	CAATTATTAC AATCTTATGT CATTAAAGG ATTCAAAAAG GGTGTCTTA GGGCCCCACA	240
	GACAATGCGT CAGAAATTCA ACATGGGAGA AATCACCCAA GATGCTGTTT ATCTCGATGC	300
	TCGAAAGAAGA TTCAAAAGAAA TCGAAACGGA AACAAAAAAG TTGAGTGAAG AATCCAAGAA	360
	ATATTCAT GCTGTCAATG GGATGTTAGA TGAACAAATT GATTTGCCA AAGCCGTGGC	420
15	TGAGATTAT AAACCAATCA GTGGTAGATT ATCGGACCCC AGTGCTACGG TACCAAGAAGA	480
	TAACCCACAA GGTATTGAAG CATCGGAAC GTACCAAGCA GTGGTAAAG ATCTCAAAGA	540
	TACCTAAAA CCCGATTGG AATTGATTGA AAAAAGAATT GTTGAACCGAG CACAAGAATT	600
20	ATTGAAGATT ATACAAGCTA TAAGGAAAAT GTCAAGTGAAG AGAGACCATA AACATTGGA	660
	TTGGGATCGT CATAAGAGAA ATTTTCTAA ATATGAAC TGAAAGAAAA GAAGTGTAA	720
	AGATGAAGAA AAAATGTTCA GTGCTCAANC AGAAGTAGAA ATTGCTAAC AAGAGTACGA	780
25	TTTATTATAAT GATTGTTAA AGAATGAATT GCCAGTTTG TTTCAAATGC AAAGTGTATT	840
	TATCAAACCA TTGTTGTTT CATTCTATTA CATGCAGTTG AATATTTCT ACACATTATA	900
	CACTAGAATG GAAGAGTTGA AAATTCCATA TTTTGATTG TCTACTGATA TTGTCGAAGC	960
30	TTATACTGCC AAGAAGGGGA ACATTGAGGA ACACCCGAT GCTATTGGAA TCACTCATT	1020
	CAAAGTCGGG CATGCCAAT CCAAATTGGA AGCCACTAAA AGAAGACATG CTGCTATGAA	1080
	TAGTCCACCT CCTACCGGTG CCAGCTCTAT TGCATCTACA GGTAATGGTG GTGAATTACC	1140
	TGCATACTCC CCAGGAGGTT ACAACCAACC ATATGGTAT AGCAAGTATC AACACCAC	1200
35	TTCTCCAGCA ACATACCAAT CTCCAGTAGT AGCAGCCACT GCTCAATCTC CAGCTACTTA	1260
	TCAATGCCA GTGGCTACTG GACAACCTCC ATCATATTAA CCACAAACTC CAGCCAGTGC	1320
	TCCACCACCA CAAGTTGGTA GTGGCCTTCC AACATGCACG GCTTATACG ATTATACTGC	1380
40	ACAAGCCAG GGTGACTTGA CTTTCCCTGC AGGAGCTGTT ATTGAATTAA TACAAAGAAC	1440
	CGAAGATGCC AACGGATGGT GGACTGGTAA ATACAATGGT CAAACGGTG TGTTCCCTGG	1500
	TAATTATGTG CAATTATAG	1519

(2) INFORMATION FOR SEQ ID NO: 66:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 440 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

50

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

55

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

Met Ser Phe Lys Gly Phe Lys Lys Gly Val Leu Arg Ala Pro Gln Thr
1 5 10 15

5 Met Arg Gln Lys Phe Asn Met Gly Glu Ile Thr Gln Asp Ala Val Tyr
20 25 30

Leu Asp Ala Glu Arg Arg Phe Lys Glu Ile Glu Thr Glu Thr Lys Lys
35 40 45

10 Leu Ser Glu Glu Ser Lys Lys Tyr Phe Asn Ala Val Asn Gly Met Leu
50 55 60

Asp Glu Gln Ile Asp Phe Ala Lys Ala Val Ala Glu Ile Tyr Lys Pro
65 70 75 80

Ile Ser Gly Arg Leu Ser Asp Pro Ser Ala Thr Val Pro Glu Asp Asn
85 90 95

15 Pro Gln Gly Ile Glu Ala Ser Glu Ser Tyr Gln Ala Val Val Lys Asp
100 105 110

Leu Lys Asp Thr Leu Lys Pro Asp Leu Glu Leu Ile Glu Lys Arg Ile
115 120 125

20 Val Glu Pro Ala Gln Glu Leu Leu Lys Ile Ile Gln Ala Ile Arg Lys
130 135 140

Met Ser Val Lys Arg Asp His Lys Gln Leu Asp Leu Asp Arg His Lys
145 150 155 160

25 Arg Asn Phe Ser Lys Tyr Glu Ser Lys Lys Glu Arg Thr Val Lys Asp
165 170 175

Glu Glu Lys Met Phe Ser Ala Gln Xaa Glu Val Glu Ile Ala Gln Gln
180 185 190

Glu Tyr Asp Tyr Tyr Asn Asp Leu Leu Lys Asn Glu Leu Pro Val Leu
195 200 205

30 Phe Gln Met Gln Ser Asp Phe Ile Lys Pro Leu Phe Val Ser Phe Tyr
210 215 220

Tyr Met Gln Leu Asn Ile Phe Tyr Thr Leu Tyr Thr Arg Met Glu Glu
225 230 235 240

35 Leu Lys Ile Pro Tyr Phe Asp Leu Ser Thr Asp Ile Val Glu Ala Tyr
245 250 255

Thr Ala Lys Lys Gly Asn Ile Glu Glu Gln Thr Asp Ala Ile Gly Ile
260 265 270

40 Thr His Phe Lys Val Gly His Ala Lys Ser Lys Leu Glu Ala Thr Lys
275 280 285

Arg Arg His Ala Ala Met Asn Ser Pro Pro Pro Thr Gly Ala Ser Ser
290 295 300

Ile Ala Ser Thr Gly Thr Gly Gly Glu Leu Pro Ala Tyr Ser Pro Gly
305 310 315 320

45 Gly Tyr Asn Gln Pro Tyr Gly Asp Ser Lys Tyr Gln Pro Pro Ser Ser
325 330 335

Pro Ala Thr Tyr Gln Ser Pro Val Val Ala Ala Thr Ala Gln Ser Pro
340 345 350

50 Ala Thr Tyr Gln Ser Pro Val Ala Thr Gly Gln Pro Pro Ser Tyr Leu
355 360 365

Pro Gln Thr Pro Ala Ser Ala Pro Pro Gln Val Gly Ser Gly Leu

EP 0 982 401 A2

370	375	380		
5	385	390	395	400
	Leu Thr Phe Pro Ala Gly Ala Val Ile Glu Ile Ile Gln Arg Thr Glu			
	405		410	415
	Asp Ala Asn Gly Trp Trp Thr Gly Lys Tyr Asn Gly Gln Thr Gly Val			
	420		425	430
10				
	Phe Pro Gly Asn Tyr Val Gln Leu			
	435		440	

(2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 855 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

25	ATAATTTCA GAAAGAGACT AGATTCTGAT AGAAAATATAG ACGCATCACT ATATTTGGA	60
	AATATAGATC CACAAGTTAC GGAGTTGTTA ATGTATGAGT TGTTCATCCA ATTTGGTCCC	120
	GTCAAATCAA TCAATATGCC AAAGGATCGT ATATTGAAAA CACACCAGGG GTATGGATT	180
	GTCGAATTAA AAAACTCAGC AGATGCCAAA TATACTATGG AAATACTACG AGGAATAAGA	240
30	CTTTATGGAA AAGCATTGAA ATTGAAACGA ATTGATGCCA AGTCTCAGTC ATCAACAAAC	300
	AACCCAAATA ATCAAACAAT AGGAACATT GTACAATCAG ATTTGATCAA TCCAAATTAC	360
	ATAGATGTTG GAGCTAAACT ATTTATCAAC AATCTTAATC CATTGGTCGA TGAATCCTT	420
	TTAATGGATA CGTTTAGTAA GTTTGGAACC CTTATAAGAA ACCCAATAAT TAGACGTGAT	480
35	TCAGAGGGAC ACTCTTGGG ATACGGATT CTTACGTACG ATGACTTTGA AAGTAGTGAT	540
	TTATGCATAC AAAAAATGAA CAACACGATT TTGATGAATA AAAAAATTGC TATCAGTTAT	600
	GCATTCAAGG ATCTGAGTGT TGATGGGAAG AAATCCCGGC ATGGAGATCA AGTGGAGCGG	660
	AAATTGGCTG AAAGTGCCAA AAAGATAAT TTGTTGGTAA CGAAAAACTTC TAAGGCAGGT	720
40	ACGACGAAGG GAAATAAAAG GAAGATAAAA CCACATAAAG TGACCAAACC GTGAGACAAT	780
	GAGTTAGCTC CCCCTTCAA AATAAGTAGA GTATCACCAT AGTTTATGAA ACAATTGATA	840
	TATTAAGCTT CTCTG	855

45

(2) INFORMATION FOR SEQ ID NO: 68:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 257 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO

55

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

Ile Ile Phe Arg Lys Arg Leu Asp Ser Asp Arg Asn Ile Asp Ala Ser
1 5 10 15

5 Leu Tyr Phe Gly Asn Ile Asp Pro Gln Val Thr Glu Leu Leu Met Tyr
20 25 30

Glu Leu Phe Ile Gln Phe Gly Pro Val Lys Ser Ile Asn Met Pro Lys
35 40 45

10 Asp Arg Ile Leu Lys Thr His Gln Gly Tyr Gly Phe Val Glu Phe Lys
50 55 60

Asn Ser Ala Asp Ala Lys Tyr Thr Met Glu Ile Leu Arg Gly Ile Arg
65 70 75 80

15 Leu Tyr Gly Lys Ala Leu Lys Leu Lys Arg Ile Asp Ala Lys Ser Gln
85 90 95

Ser Ser Thr Asn Asn Pro Asn Asn Gln Thr Ile Gly Thr Phe Val Gln
100 105 110

20 Ser Asp Leu Ile Asn Pro Asn Tyr Ile Asp Val Gly Ala Lys Leu Phe
115 120 125

Ile Asn Asn Leu Asn Pro Leu Val Asp Glu Ser Phe Leu Met Asp Thr
130 135 140

25 Phe Ser Lys Phe Gly Thr Leu Ile Arg Asn Pro Ile Ile Arg Arg Asp
145 150 155 160

Ser Glu Gly His Ser Leu Gly Tyr Gly Phe Leu Thr Tyr Asp Asp Phe
165 170 175

30 Glu Ser Ser Asp Leu Cys Ile Gln Lys Met Asn Asn Thr Ile Leu Met
180 185 190

Asn Asn Lys Ile Ala Ile Ser Tyr Ala Phe Lys Asp Ser Ser Val Asp
195 200 205

35 Gly Lys Lys Ser Arg His Gly Asp Gln Val Glu Arg Lys Leu Ala Glu
210 215 220

Ser Ala Lys Lys Asn Asn Leu Leu Val Thr Lys Thr Ser Lys Ala Gly
225 230 235 240

40 Thr Thr Lys Gly Asn Lys Arg Lys Asn Lys Pro His Lys Val Thr Lys
245 250 255

Pro

(2) INFORMATION FOR SEQ ID NO: 69:

45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

5	CTGTTTATTAAATGGATATA TGTTAACCA TGAACCTTCGG TTTATCAGAA AAATTGGTGC	60
	TGGTACCTAT GGTTTGATTT ACCTTGTGGA AAATATCTAC ACTAAACAAAC AATTTGCTGC	120
	TAAAATGGT CTTGAACAGC CATTACTCAA ACAAAAGCAA CAACAACAAAC AAAGTCATCA	180
	TGGACATAAA GGAGAATCTA GTATGAACAA ACAAAATAATA CTGCAAGAAT TTTATCAATA	240
10	TTTTTTAAC AATAGTATGC CACAACCCAG AAATTGGAC TTGAATTACCTTCGAGACAA	300
	CGGACATGAT TGCCCCTTTT TGACTGAAAT CTCATTACAT TTAAAGTAC ATCAACACCC	360
	AAACATAGCG ACTATTCACTC AAGTATTAAA CATTGAAGAT TTTGCCATAA TAATATTGAT	420
	GGATCATTG GAGCAAGGGG ATTTGTCAC TAATATCATT GATAGACAAA TATTCAACAA	480
15	TAATAGTCAT AGAAAAGTTC CAAGAACAGA TTTGAAACCC CAATTATTAA TGAAGAATGC	540
	CATTTACAA TTGATAGAAG CCATTGAATA TTGTCACCGAA AATAATATTT ACCATTGTGA	600
	TTTAAACCA GAAAACATTA TGGTTAGATA TAATCCATAC TATGTTCGTC CAACTATCAA	660
	TAACAATAAT AACAATGGGAG AAGATGATTG ATGCTATGCC AACAGTATTAA TTGACTATAA	720
20	TGAATTACAC CTCGTGTTGA TTGATTGGG TTTAGCTATG GACTCTGCTA CCATTGTTG	780
	TAATTCACTGT CGTGGATCGT CATTTCATAC TATGTTCGTC CAACTATCAA	840
	CCATCGTTA ATCAACCAAT TAATTGATAT GAATCAATAT GAGTCATTG AAATCAATGG	900
	GACAACAGTG ACAAAATCAA ACTGTAAATAA TTTACCTACA TTGGCTGGGG ATATTGGTC	960
25	ATTGGGGATA TTGTTCTTAA ATATCACTTG TTCAAGAAC CCATGGCCCA TTGCATCATT	1020
	TGATAATAAT CAAAATAATG AAGTGTAAAGAATATGTTAA GAATTATATG TTGAAATAATA ACAAGGCTGT	1080
	TTTGAGCAAA ATCTTACCA TTTCTCACA ATTTAATCGC TTATTAGATA GAATTTCAA	1140
30	ATTGAATCCT AATGATAGAA TAGATTACCA AACTTTATAC AAAGAAGTTA TTGTTGTGA	1200
	TTTCTCAAA GATGATCATT ACTACTATGC CCAACATCAA CATCATCACA ATCACATCA	1260
	AATCAATAAT GCTTACAATC ACTATCAGAA ACAACCTAAT CAAGCAAGAC CTACTGCAAA	1320
	CCAACAATTG TATACACCAC CGGAAACCCAC CACTTATAAT TCATACGCTA GTGATATGGA	1380
35	AGAAGATGAA ATTAGTGTGATG ATGAGTTTA TTCTGATGAA GAAGATGAAG ATATTGAAGA	1440
	CTATGAAGAG GAAGAGGAAG AGTATTTGG TAATGAGCAA CAACAACAAAC AGCAAGTCAC	1500
	AACAGTGAAT GGTAATTTTG GTCAAGTTAA AGGTACCTGT TATTACGATA CCAAAACCAA	1560
40	AACAACATACA TATATAAAAC CACCAAGCTGC ATATACTTTA GAGACGCCCTA GTCAAAGTGT	1620
	TGAATACTGT TAAGTTGTAC ACATAAATAA TTAATGACAA TTAATAATAA CGATTAATAA	1680
	TATAG	1685

45 (2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 537 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

50 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

EP 0 982 401 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

Met Leu Asn His Glu Leu Arg Phe Ile Arg Lys Ile Gly Ala Gly Thr
 1 5 10 15

5 Tyr Gly Leu Ile Tyr Leu Val Glu Asn Ile Tyr Thr Lys Gln Gln Phe
 20 25 30

Ala Ala Lys Met Val Leu Glu Gln Pro Leu Leu Lys Gln Lys Gln Gln
 35 40 45

10 Gln Gln Gln Ser His His Gly His Lys Gly Glu Ser Ser Met Asn Lys
 50 55 60

Gln Ile Ile Ser Gln Glu Phe Tyr Gln Tyr Phe Leu Asn Asn Ser Met
 65 70 75 80

15 Pro Gln Pro Arg Asn Leu Asp Leu Asn Tyr Leu Arg Asp Asn Gly His
 85 90 95

Asp Cys Pro Phe Leu Thr Glu Ile Ser Leu His Leu Lys Val His Gln
 100 105 110

His Pro Asn Ile Ala Thr Ile His Gln Val Leu Asn Ile Glu Asp Phe
 115 120 125

20 Ala Ile Ile Ile Leu Met Asp His Phe Glu Gln Gly Asp Leu Phe Thr
 130 135 140

Asn Ile Ile Asp Arg Gln Ile Phe Thr Asn Asn Ser His Arg Lys Val
 145 150 155 160

25 Pro Arg Thr Asp Phe Glu Thr Gln Leu Leu Met Lys Asn Ala Met Leu
 165 170 175

Gln Leu Ile Glu Ala Ile Glu Tyr Cys His Glu Asn Asn Ile Tyr His
 180 185 190

30 Cys Asp Leu Lys Pro Glu Asn Ile Met Val Arg Tyr Asn Pro Tyr Tyr
 195 200 205

Val Arg Pro Thr Ile Asn Asn Asn Asn Asn Gly Glu Asp Asp Leu
 210 215 220

35 Cys Tyr Ala Asn Ser Ile Ile Asp Tyr Asn Glu Leu His Leu Val Leu
 225 230 235 240

Ile Asp Phe Gly Leu Ala Met Asp Ser Ala Thr Ile Cys Cys Asn Ser
 245 250 255

Cys Arg Gly Ser Ser Phe Tyr Met Ala Pro Glu Arg Thr Thr Asn Tyr
 260 265 270

40 Asn Thr His Arg Leu Ile Asn Gln Leu Ile Asp Met Asn Gln Tyr Glu
 275 280 285

Ser Ile Glu Ile Asn Gly Thr Thr Val Thr Lys Ser Asn Cys Lys Tyr
 290 295 300

45 Leu Pro Thr Leu Ala Gly Asp Ile Trp Ser Leu Gly Val Leu Phe Ile
 305 310 315 320

Asn Ile Thr Cys Ser Arg Asn Pro Trp Pro Ile Ala Ser Phe Asp Asn
 325 330 335

50 Asn Gln Asn Asn Glu Val Phe Lys Asn Tyr Met Leu Asn Asn Asn Lys
 340 345 350

Ala Val Leu Ser Lys Ile Leu Pro Ile Ser Ser Gln Phe Asn Arg Leu

EP 0 982 401 A2

	355	360	365	
	Leu Asp Arg Ile Phe Lys	Leu Asn Pro Asn Asp Arg	Ile Asp Leu Pro	
	370	375	380	
5	Thr Leu Tyr Lys Glu Val Ile Arg Cys Asp Phe Phe Lys Asp Asp His			
	385	390	395	400
	Tyr Tyr Tyr Ala Gln His Gln His His Asn His Asn Gln Ile Asn			
	405	410	415	
10	Asn Ala Tyr Asn His Tyr Gln Lys Gln Pro Asn Gln Ala Arg Pro Thr			
	420	425	430	
	Ala Asn Gln Gln Leu Tyr Thr Pro Pro Glu Thr Thr Tyr Asn Ser			
	435	440	445	
15	Tyr Ala Ser Asp Met Glu Glu Asp Glu Ile Ser Asp Asp Glu Phe Tyr			
	450	455	460	
	Ser Asp Glu Glu Asp Glu Asp Ile Glu Asp Tyr Glu Glu Glu Glu			
	465	470	475	480
	Glu Tyr Phe Gly Asn Glu Gln Gln Gln Gln Gln Val Thr Thr Val			
	485	490	495	
20	Asn Gly Asn Phe Gly Gln Val Lys Gly Thr Cys Tyr Tyr Asp Thr Lys			
	500	505	510	
	Thr Lys Thr Thr Thr Tyr Ile Lys Pro Pro Ala Ala Tyr Thr Leu Glu			
	515	520	525	
25	Thr Pro Ser Gln Ser Val Glu Tyr Cys			
	530	535		

(2) INFORMATION FOR SEQ ID NO: 71:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 848 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

40	AACCAATTTC AGAAAACAATG GCTCGTCAAT TTTTCGTAAGG TGGTAACTTC AAGGCTAACG	60
	GTACCAAAACA ACAAAATCACT TCAATCATCG ACAAACTGAA CAAGGCTGAT TTACCAAAGG	120
	ATGTCGAAGT TGTCAATTGT CCACCCGCC CTTACCTTGG TTTAGCTGTA GAGCAAAACA	180
	AACAAACCAAC TGGTGCATT GGTGCTAAA ATGTTTTGA CAACTCATGT GGTGCTTCA	240
45	CTGGTGAAAC CTGTGCTTCT CAAATCTTGG ATGTTGGTGC CAGCTGGACT TTAACTGGTC	300
	ACAGTGAAG AAGAACCATT ATCAAAGAAT CCGATGAATT CATTGCTGAA AAAACCAAGT	360
	TTGCCTTGGGA CACTGGTGTCA AAAGTTATT TATGTATTGG TGAAACCTTA GAGGAAAGAA	420
50	AAGGTGGTGT CACTTTGGAT GTTGTGCCA GACAATTGGA TGCTGTTCC AAGATTGTT	480
	CTGATTGGTC AACACATTGTT GTTGCTTACCG AACCTGTTG GCCAATTGGT ACTGGTTAG	540
	CCGCTACCCC AGAAGATGCT GAAGAAACCC ACAAAAGGTAT TAGAGCTCAT TTGGCCAAGA	600

55

EP 0 982 401 A2

5	CCATGGTGC CGAACAAAGCT GAAAAAACCA GAATCTTGTCA CGGTGGTCA GTTAACGGTA	660
	AGAACGCTAA GGATTCAAA GACAAAGCAA ATGTTGATGG TTTCTTAGTC GGTGGTGCTT	720
	CATTTAAACC AGAATTGTT GATATCATCA AATCTAGATT ATAAACAGTA TATTAAAAAC	780
	TATATGCCTA TAGAATTTAG CATGTTGTTG TGAATTTGTA ATGAATCTAT AAAAATGTGC	840
	TCATGAAC	848

(2) INFORMATION FOR SEQ ID NO: 72:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 248 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

15 (ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:
 Met Ala Arg Gln Phe Phe Val Gly Gly Asn Phe Lys Ala Asn Gly Thr
 1 5 10 15
 Lys Gln Gln Ile Thr Ser Ile Ile Asp Asn Leu Asn Lys Ala Asp Leu
 20 25 30
 25 Pro Lys Asp Val Glu Val Val Ile Cys Pro Pro Ala Leu Tyr Leu Gly
 35 40 45
 Leu Ala Val Glu Gln Asn Lys Gln Pro Thr Val Ala Ile Gly Ala Gln
 50 55 60
 30 Asn Val Phe Asp Lys Ser Cys Gly Ala Phe Thr Gly Glu Thr Cys Ala
 65 70 75 80
 Ser Gln Ile Leu Asp Val Gly Ala Ser Trp Thr Leu Thr Gly His Ser
 85 90 95
 Glu Arg Arg Thr Ile Ile Lys Glu Ser Asp Glu Phe Ile Ala Glu Lys
 100 105 110
 35 Thr Lys Phe Ala Leu Asp Thr Gly Val Lys Val Ile Leu Cys Ile Gly
 115 120 125
 Glu Thr Leu Glu Glu Arg Lys Gly Gly Val Thr Leu Asp Val Cys Ala
 130 135 140
 40 Arg Gln Leu Asp Ala Val Ser Lys Ile Val Ser Asp Trp Ser Asn Ile
 145 150 155 160
 Val Val Ala Tyr Glu Pro Val Trp Ala Ile Gly Thr Gly Leu Ala Ala
 165 170 175
 45 Thr Pro Glu Asp Ala Glu Glu Thr His Lys Gly Ile Arg Ala His Leu
 180 185 190
 Ala Lys Thr Ile Gly Ala Glu Gln Ala Glu Lys Thr Arg Ile Leu Tyr
 195 200 205
 50 Gly Gly Ser Val Asn Gly Lys Asn Ala Lys Asp Phe Lys Asp Lys Ala
 210 215 220
 Asn Val Asp Gly Phe Leu Val Gly Ala Ser Leu Lys Pro Glu Phe
 225 230 235 240

Val Asp Ile Ile Lys Ser Arg Leu
245

5

Claims

1. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides in Sequence ID Numbers 1 to 10 3, 5, 6, 8 to 11, 13, 15, 16, 18, 20, 21, 23, 25 to 29, 31, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69 and 71.

2. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides in Sequence ID Numbers 28, 35, 37 and 39 and fragments or derivatives of said nucleic acid molecules.

15 3. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which polypeptide has an amino acid sequence according to the sequence of any of Sequence ID Numbers 4, 7, 12, 14, 17, 19, 22, 24, 30, 32 to 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 and 72.

20 4. A nucleic acid molecule according to any of claims 1 to 3 which is mRNA.

5. A nucleic acid molecule according to any of claims 1 to 3 which is DNA.

6. A nucleic acid molecule according to claim 5 which is cDNA.

25 7. A nucleic acid molecule capable of hybridising to the molecules according to any of claims 1 to 5 under high stringency conditions.

8. A polypeptide having the amino acid sequences of any of Sequence ID Numbers 4, 7, 12, 14, 17, 19, 22, 24, 30, 32 to 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 and 72.

30 9. A polypeptide encoded by the nucleic acid molecule according to any of claims 1 to 6.

10. A polypeptide according to claim 9 having an amino acid sequence of any of Sequence ID Numbers 4, 7, 12, 14, 17, 19, 22, 24, 30, 32 to 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 and 72.

35 11. An expression vector comprising a nucleic acid molecule according to claim 5 or 6.

12. An expression vector according to claim 11 which comprises an inducible promoter.

13. An expression vector according to claim 11 or 12 which comprises a sequence encoding a reporter molecule.

14. A nucleic acid molecule according to any of claims 1 to 7 for use as a medicament.

40 15. Use of a nucleic acid molecule according to any of claims 1 to 7 in the preparation of a medicament for treating *Candida albicans* associated diseases.

16. A polypeptide according to any of claims 8 or 10 for use as a medicament.

45 17. Use of a polypeptide according to any of claims 8 to 10 in the preparation of a medicament for treating *Candida albicans* associated infections.

18. A pharmaceutical composition comprising a nucleic acid molecule according to any of claims 1 to 7 or a polypeptide according to any of claims 8 to 10 together with a pharmaceutically acceptable carrier diluent or excipient therefor.

50 19. A *Candida albicans* cell comprising an induced mutation in the DNA sequence encoding the polypeptide according to any of claims 8 to 10.

20. A method of identifying compounds which selectively modulate expression of polypeptides which are crucial for 55 growth and survival of *Candida albicans*, which method comprises:

(a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to any of claims 1 to 6 which mutation results in overexpression or

EP 0 982 401 A2

underexpression of said polypeptides in addition to contacting one or more wild type *Candida albicans* cells with said compound,

(b) monitoring the growth and/or activity of said mutated cell compared to said wild type; wherein differential growth or activity of said one or more mutated *Candida* cells is indicative of selective action of said compound on a polypeptide or another polypeptide in the same or a parallel pathway.

- 5 21. A compound identifiable according to the method of claim 20.
- 10 22. A compound according to claim 21 for use as a medicament.
23. Use of a compound according to claim 21 in the preparation of a medicament for treating *Candida albicans* associated diseases.
- 15 24. A pharmaceutical composition comprising a compound according to claim 21 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.
- 25 25. A method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival of said cell or organism, which method comprises:
 - 20 (a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide sequences in said cDNA or genomic library,
 - (b) selecting transformants exhibiting impaired growth and determining the nucleotide sequence of the cDNA or genomic sequence from the library included in the vector from said transformant.
26. A method according to claim 25 wherein said cell or organism is a yeast or filamentous fungi.
27. A method according to claim 25 or 26 wherein said cell or organism is any of *Saccharomyces cerevisiae*, *Saccharomyces pombe* or *Candida albicans*.
- 30 28. Plasmid pGAL1PSiST-1 having the sequence of nucleotides illustrated in Figure 2.
29. Plasmid pGAL1PNiST-1 having the sequence of nucleotides illustrated in Figure 4.
- 30 30. An antibody capable of binding to a polypeptide according to any of claims 8 or 10.
- 35 31. An oligonucleotide comprising a fragment of from 10 to 50 contiguous nucleic acid sequences of a nucleic acid molecule according to any of claims 1 to 7.
32. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans*, said nucleic acid molecule comprising the sequences of any of the nucleotide sequences illustrated in Figures 5 to 28.
- 40 33. A polypeptide which is critical for survival and growth of the yeast *Candida albicans*, said polypeptide comprising the amino acid sequences of any of the sequences illustrated in Figures 29 to 39.

45

50

55

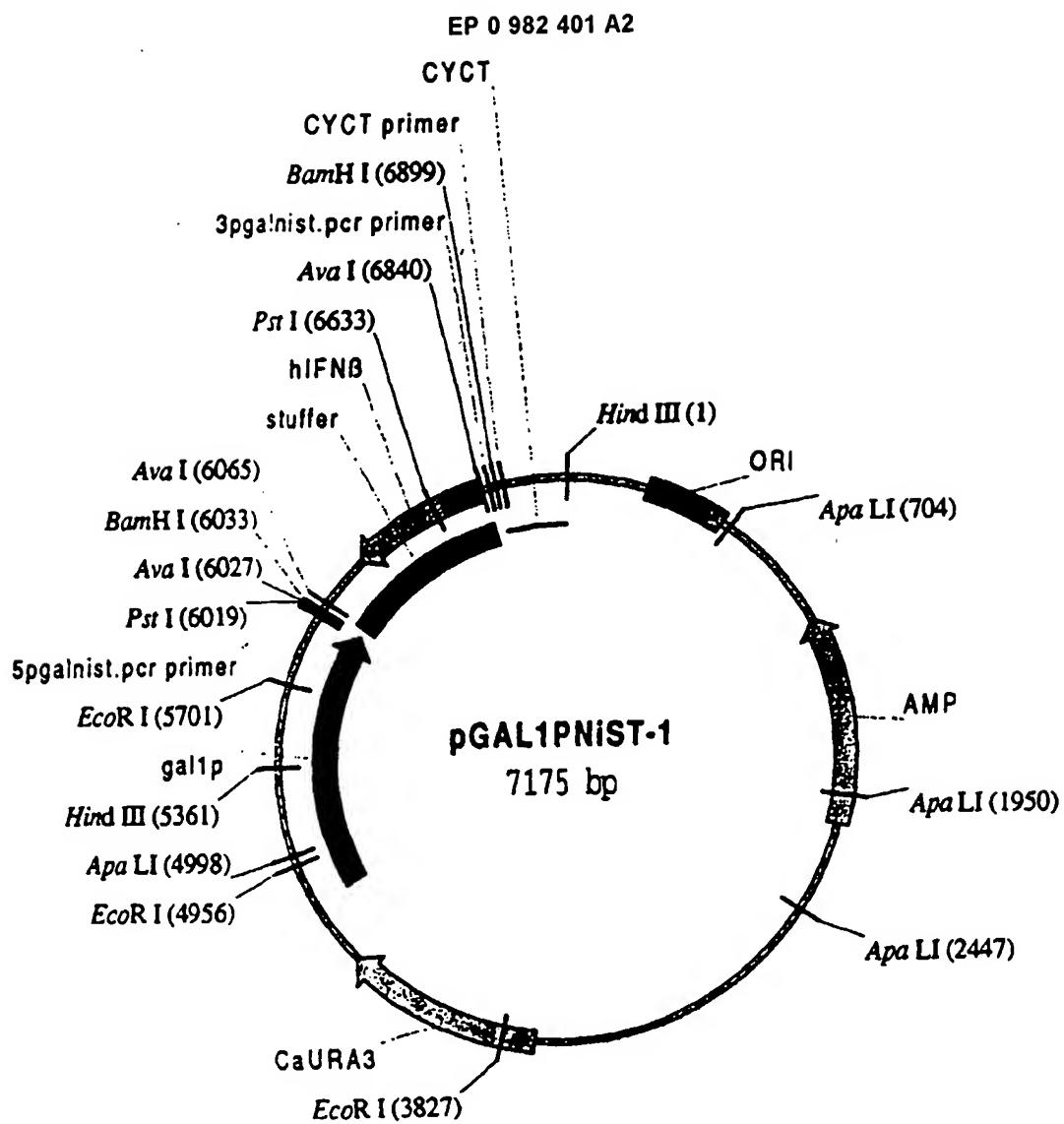


Figure 1

HindIII

1 AGCTTGAGTA TTCTATAGTG TCACCTAAAT AGCTTGGGT AATCATGGTC
 TCGAACTCAT AAGATATCAC AGTGGATTA TCGAACCGCA TTAGTACCG

 51 ATAGCTGTCTT CCTGTGTGAA ATGTTATCC GCTCACAAATT CCACACAAAC
 TATCGACAAA GGACACACTT TAACAAATAGG CGAGTGTAA GGTGTTGT

 101 TACGAGCCG AAGCATAAAG TGTAAAGCCT GGGGTGCTTA ATGAGTGAGC
 ATGCTGGCC TTCTATTC ACATTCGGA CCCACGGAT TACTCACTCG

 151 TAACTCACAT TAATTGGTTT GCGCTCACTG CCCGCTTTCC AGTCGGAAA
 ATTGAGTGTA ATTAAGCAA CCCGAGTGAC GGGGAAAGG TCAGCCCTTT

 201 CCTGTGTGCA CACCTGCATT AATGAATCGG CCACCGCGG GGGAGAGCG
 GGACAGCAAG GTGAGCTAA TTACTTAGCC GTTGGCGGC CCCTCTCCGC

 251 GTTTCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTGGCTGGC
 CAAACGCATA ACCCGCGAGA AGCGGAAGGA GCGAGTGACT GAGGGACCG

 301 TCGGTCGTTG GGCTGGCGG AGCGGTATCA GCTCACTCAA AGGGGTAAT
 AGCCAGCAAG CCGACGGCGC TCGCCATAGT CGAGTGTAGT TCGCCATTA

 351 ACGGTTATCC ACAGAACATCG GGGATAACGC AGGAAAGAAC ATGTGAGCAA
 TCCAATAGG TGTCTTAGTC CCCTATTGCG TCTTTCTTG TACACTCGTT

 401 AAGGCCACCA AAAGGCCAGG AACCGTAAAA AGGCCCGTT CCTGGGTTT
 TTCCGGTCTT TTCCGGTCC TTGGCATTT TCCGGCCAA CGACCCCAA

 451 TTCCATAGGC TCCGCCCCC TGACGAGCAT CACAAAATC GACGGCTCAAG
 AAGGTATCGG AGGGGGGGG ACTGCTGTA GTGTTTTAG CTGGAGGTT

 501 TCAAGGGTGG CGAAACCCGA CAGGACTATA AAGATACCG AGGTTTCCCG
 AGTCTCCACC GCTTGGGT GTCTGTATAT TTCTATGTC CGCAAAAGGG

 551 CTGGAAGTC CTCGTGGC TCTCTGTTC CGACCCCTGCC GCTTACCGGA
 GACCTTCGAG GGAGCACCG AGAGGACAAG GCTGGGACGG CGAATGGCCT

 601 TACCTGTCCG CCTTTCTCCC TTGGGGAGG GTGGCGCTT CTCATAGCTC
 ATGGACAGGC GGAAAGAGGG AAGCCCTTCG CACCGGAAA GAGTATCGAG

 651 ACGCTGTAGG TATCTCAAGT CGGTGTAGGT CGTTGCTCC AAGCTGGCT
 TGCGACATCC ATAGAGTCAA CCCACATCCA GCAAGCGAGG TTGGACCCGA

ApaLI

701 GTGTGCACGA ACCCCCCCTT CAGCCCGACC GCTGGCCCTT ATCCGGTAAC
 CACACGTGCT TGGGGGCGA GTGGGGCTGG CGACCGGGAA TAGGCCATTG

 751 TATCGTCTTG AGTCCAACCC GGTAAAGACAC GACTTATCGC CACTGGCAGC
 ATAGCAGAAC TCAGGTTGGG CCATTCTGT CTGAATAGCG GTGACCGTGC

 801 AGCCACTGGT AACAGGATTA CGAGAGCGAG GTATGTAGGC GGTGCTACAG
 TCGGTGACCA TTGTCTTAAT CGTCTCGCTC CATAACATCGG CCACGATGTC

 851 AGTCTTCAAA GTGGTGGCCTT AACTACGGCTT AACTAGAAAG GACAGTATTT
 TCAAGAACTT CACCACCGGA TTGATGCCA TGTGATCTTC CTGTCTAAA

 901 GGTATCTGCG CTCTGCTGAA CCCAGTTACC TTGGAAAAAA GAGTTGGTAG
 CCATAGACCC GAGACGACTT CGGTCAATGG AAGCTTTTT CTCAACCATC

Fig.2

951 CTC TTGATCC GGCAACAAA CCACCGCTGG TAGGGGTGGT TTTTTTGTGTT
 GAGAACTAGG CGGTTGTTT GGTGGCGACC ATGGCCACCA AAAAAACAAA

 1001 GCAAGCAGCA GATTACGCCG AGAAAAAAAG GATCTCAAGA AGATCCTTGG
 CGTTCTCGT CTAATGCGCG TCTTTTTTC CTAGAGTTCT TCTAGGAAC

 1051 ATCTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAAGT CACGTTAAGG
 TAGAAAAGAT GCCCCAGACT GCGAGTCACC TTGCTTTGAGA GTGCAATTCC

 1101 GATTTGGTC ATGAGATTAT CAAAAGGAT CTTCACCTAG ATCCTTTAA
 CTAAAACAG TACTCTAATA GTTTTCTA GAAGTGGATC TAGGAAAATT

 1151 ATTTAAAATC AAGTTTAA TCAATCTAAA GTATATATGA GTAAACTTGG
 TAATTTTAC TTCAAAATT AGTTAGATTT CATATATACT CATTGAACC

 1201 TCTGACAGTT ACCAATGCTT AATCACTGAG GCACCTATCT CAGCGATCTG
 AGACTCTAA TGGTTACGAA TTACTCACTC CGTGGATAGA GTGCGTAGAC

 1251 TCTATTCTGTCATAG TTGCTGACT CCCCGCTGG TAGATAACTA
 AGATAAAGCA AGTAGGTATC AACGGACTGA GGGCAGCAC ATCTATTGAT

 1301 CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACGGGA
 GCTATGCCCT CCCGAATGGT AGACGGGGT CACGACGTTA CTATGGCCCT

 1351 GACCCACGCT CACCGGCTCC AGATTATCA GCAATAAACCG AGCCAGCCGG
 CTGGGTCCGA GTGGCCGAGG TCTAAATAGT CGTTATTTGG TCGGTGGCC

 1401 AAGGGCCGAG CGCAGAAGTG GTCTGCAAC TTATCCGCC TCCATCCAGT
 TTCCCGCTC GCGTCTTCAC CAGGACGTTG AAATAGGCC AGGTAGGTCA

 1451 CTATTAATTG TTGCGGGAA GCTAGAGTAA GTAGTTGCC AGTTAATAGT
 GATAATTAAAC AACGGCCCTT CGATCTCATT CATCAAGCGG TCAATTATCA

 1501 TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC
 AACGGGTGCA AACAACGGTA ACGATGTCGG TAGCACCACA GTGGGAGCAG

 1551 GTTTGGTATG GCTTCATTCAC GCTCCGGTTC CCAACGATCA AGGGAGTTA
 CAAACCATAC CGAAGTAAGT CGAGGCCAAG GGTGCTAGT TCGCTCAAT

 1601 CATGATCCCC CATGTTGTC AAAAAGCGG TTAGCTCCCTT CGGTCCCTCCG
 GTACTAGGG GTACAACACG TTTTTGCC AATCGAGGAA GCCAGGAGCC

 1651 ATCGTTGTCAC GAACTAAGT GGCCTCAGTG TTATCACTCA TGTTTATGCC
 TAGCAACAGT CTTCATTCAC CCGCGTCAC AATAGTGAGT ACCAATACCG

 1701 AGCACTGCAT AATTCTCTA CTGTCATGCC ATCGTAAGA TGCTTTCTG
 TCGTGACGTA TAAAGAGAAT GACAGTACGG TAGGATTCTT AGAAAAGAC

 1751 TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAAATAGTG TATGCGGGGA
 ACTGACCACT CATGAGTTGG TTCAGTAAGA CTCTTATCAC ATACGGCGCT

 1801 CCGACTGGT CTTCCCCGGC GTCAATACGG GATAATACCG CCCCACATAG
 GGCTCAACGA GAACGGGGCAGTTATGCC CTATTATGCC GCGGTGTATC

 1851 CAGAACTTAA AAGTGTCA TCATTGGAA ACGTTCTTCG GGGCGAAAAC
 GTCTTGAAT TTTCACGAGT AGTAACCTTT TGCAAGAAC CCCGCTTTG

ApaLI

1901 TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT
AGAGTTCTA GAATGGCAGC AACTCTAGGT CAAAGCTACAT TGGGTGAGCA

ApaLI

1951 GCACCCAACG GATCTTCAGC ATCTTTACT TTACCCACCG TTTCTGGTG
CGTGGGTGTA CTAGAAGTCG TAGAAAATGA AAGTGGTCG AAAGACCCAC

2001 AGCAAAACAA GGAAGGCAAATGCGCAAAG AAAGGGAAATA AGGGGCACAC
TCGTTTGTG CTTTCCGTT TACGGCGTTT TTCCCTTAT TCCCGCTGTG

2051 GGAAATGTTG AATACTCATA CTCTCCCTT TTCAATATTA TTGAACCAATT
CCTTACACAC TTATGAGTAT GAGAAGGAAA AAGTTATAAT AACTTGTAA

2101 TATCAGGGTT ATTGCTCTAT GAGCGGATAC ATATTTGAAT GTATTTAGAA
ATAGTCCCAA TAACAGAGTA CTCGCTATG TATAAACTTA CATAAACTTT

2151 AAATAAAACAA ATAGGGGTTG CGGCCACATT TCCCCGAAAAA GTGCCACCTG
TTTATTGTT TATCCCAAG GCGCGTGTAA AGGGGCTTTT CACGGTGGAC

2201 ACGTCTAAGA AACCATTATT ATCATGACAT TAACCTATAA AAATAGGGT
TGCAGATTCT TTGGTAATAA TAGTACTGTA ATTGGATATT TTATCCGCA

2251 ATCACGAGGC CCTTTCTCT CGCGCGTTTC GGTGATGACG GTGAAAACCT
TAGTGCTCCG GGAAACAGA GCGCGCAAAG CCAACTACTGC CACTTTGGAA

2301 CTGACACATG CAGCTCCCG AGACGGTCAC AGCTTGTCTG TAACGGGATG
GACTGTGTAC GTCGAOGGCC TCTGCCAGTG TCGAACAGAC ATTCCCTAC

2351 CCGGGGACAG ACAAGCCCGT CAGGGCGCGT CAGGGGGTGT TGCCGGGTGT
GGCCCTCGTC TGTTCGGCA GTCCCCGCA GTCCCCACAA ACCCCCCACA

ApaLI

2401 CCGGGCTGGC TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA
GCCCGACCG AATTGATACG CCGTAGTCTC GTCTAACATG ACTCTCACGT

ApaLI

2451 CCATATGGGG TGTGAAATAC CGCACAGATG CGTAAGGAGA AAATACCGCA
GGTATACGCC ACACTTTATG GCGTGTCTAC GCATTCTCT TTTATGGCGT

2501 TCAGGGAAA TTGTAACCGT TAATATTTG TTAAAATTG CGTTAAATAT
AGTCCGCTTT AACATTGCA ATTATAAAAC AATTAAAGC GCAATTATA

2551 TTGTTAAATC AGCTCATTTT TTAACCAATA GGCGAAATC GGCAAAATCC
AACATTAG TCGAGTAAA AATTGGTTAT CCGGTTTAGG CCGTTTTAGG

2601 CTTATAAAATC AAAAGAATAG ACCGAGATAG GGTGAGTGT TGTCCAGTT
GAATATTAG TTTCTTATC TGGCTCTATC CCAACTCACA ACAAGGTCAA

2651 TGGAACAAAGA GTCCCACTATT AAAGAACGTG GACTCCAACG TCAAAGGGCG
ACCTTGTCT CAGGTGATAA TTCTTGCAC CTGAGGGTGC AGTTTCCCGC

2701 AAAAACCGTC TATCAGGGCG ATGGCCCACG ACGTGAACCA TCACCCAAAT
TTTTGCCAG ATAGTCCCGC TACCGGTGA TGCACTTGGT AGTGGGTTA

2751 CAAGTTTTT GCGGTGAGG TGCCGAAAG CTCTAAATCG GAAACCTAAA
GTTCAAAAAA CGCCAGCTCC ACGGCATTTC GAGATTAGC TTGGGATTT

EP 0 982 401 A2

2801 GGGAGCCCCC GATTTAGACC TTGACGGGGA AAGCCGGGGA ACCTGGGGAG
 CCCTCGGGGG CTAATCTCG AACTGCCCC TTGGCCGCT TGACCCGCTC

 2851 AAAGGAAGGG AAGAAAGGGA AAGGAGGGGG CGCTAGGGGG CTGGCAAGTG
 TTTCCTTCCC TTCTTCCGT TTCTCGCC GCGATCCCGC GACCGTTAAC

 2901 TAGCGGTACAC GCTGGCGGTA ACCACCAACAC CGGGCGCGCT TAATGCGCG
 ATCGCCAGTG CGACGGCAT TGGTGGTGTG GGGGGGGA ATTACGGGGC

 2951 CTACAGGGCG CGTCATTCG CCATTCAGGG TGCGCAACTG TTGGGAAGGG
 GATGTCCTCGC CGAGTAAGC GGTAACTCCG ACGGCTTGAC AACCCCTTCCC

 3001 CGATCGGTGC GGGCTCTTC GCTATTACGGC CAGCTGGCGA AAGGGGGATG
 GCTAGCCACG CCCGGAGAAG CGATAATGCG GTCGACCGCT TTCCCCCTAC

 3051 TGCTGCAAGG CGATTAAGTT GGGTAACGCC AGGGTTTCC CAGTCACCGAC
 ACAGACGTTCC GCTAATTCAA CCCATTGCGG TCCCAAAAGG GTCACTGCTG

 3101 GTTGTAACAC GACGGCCAGT GAATTGTAAT ACCGACTCACT ATAGGGCGAA
 CAACATTTG CTGGGGTCA CTTAACATTA TGCTGAGTGAA TATCCCGTT

 3151 TTGGTTTCC AATGATGAGC ACTTTAAAG TTCTGCTATG TGCGCGGTA
 AACCAAAGG TTACTACTCG TGAAAATTTC AAGACGATAC ACCGGGCCAT

 3201 TTATCCCGTG TTGACGCCGG CCAAGAGCAA CTGGTCGCC GCATACACTA
 AATAGGGCAC AACTGCGCC CGTTCTCGTT GAGCCAGCGG CGTATGTTGAT

 3251 TTCTCAAGAT GACTTGGTGT AGTACTAATAA GGAATTGATT TGGATGGTAT
 AAGACTCTTA CTGACCAAC TCATGATTAT CCTTAACCTAA ACCTACCA

 3301 AAACGAAAC AAAAAAAAGA GCTGGTACTA CTTCTTTAA AATTATTTA
 TTGCTTTG TTTTTTTCT CGACCATGAT GAAAGAATT TTAATAAAT

 3351 TTATTTGATT TTATTTATAA GTATATATTA TATTTTGAAAC GTAGATTATT
 AATAAACTAA AATAAATTAT CATATATAAT ATAAAACCTG CTTCTAATAAA

 3401 TTGTTGAAAG TTGCTGTAGT GCCATGATT CGTAACACTA ATTCTGTATT
 AACAACTTTC ACGACATCA CGGTAACTAA GCATGTTGAT TAAGACATAA

 3451 AGTCATTCCT TTGTTTGAT AGTATCAGA AAAACGGCTA TTTTTTGCA
 TCAGTAAGG GAACAAACTA TCATAGTTT TTTTGGCGAT AAAAAAACGT

 3501 ATCTTATTTTC CTGCATATTA TACACATAAC ATATGAAAG AAAAAATCTT
 TAGAATAAAG GACGTATAAT ATGCTTATTG TATTAATTTC TTTTTAGAA

 3551 TTMTTTGTT CTTCATGAT GATTCAACC ATTCTTTAA ACATTTGATCA
 AAAAAAAACAA GAAGTTACTA CTAAAGTTGG TAAGAAAATT TGTAACTAGT

 3601 ATTCCGTGAC ACAACCCCCA TACACACTGG TTATATACCC GCCCCTTTA
 TAAGGACTCG TTGTTGGGT ATGTTGACCC AAATATATGG CGGGGAAAT

 3651 CAGTTGAGA AAGAAATAGA AATAGAAATA GCAACACAAA GATATGACAG
 GTCAACTTCT TTCTTATCT TTATCTTAT CGTTGTTTCT CTAACTGTGTC

 3701 TCAACACTAA GACCTATAGT GAGAGAGCGAG AAACCTCATGC CTACCCAGTA
 AGTTGTTGATT CTGGATATCA CTCTCTCGTC TTGAGTACG GAGTGGTCAT

 3751 GCACAGCGAT TATTTGATT AATGGAACCTG AAGAAAACCA ATTATGTC
 CGTGTGCTA ATAAAGCTAA TTACCTTGAC TTCTTTGGT TAAATACACG

EcoRI

3801 ATCAATTGAC GTTGATACCA CTAAGGAATT CCTTGATTA ATTGATAAT
 TAGTTAACATG CAACTATGGT GATTCTTAA GGAACCTTAAT TAACCTATTTA

 3851 TAGGTCCCTTA TGTATGCTTA ATCAAGACTC ATATGATAT AATCAATGAT
 ATCCAGGAAT ACATACGAAT TAGTTCTGAG TATAACTATA TTAGTTACTA

 3901 TTTTCCATG AATCCACTAT TGAACCATTA TTAGAACTTT CACGTAAACA
 AAAAGGATAC TTAGGTGATA ACTMGGTAAT AATCTGAAA GTGCATTGT

 3951 TCAATTATG ATTTTGAAAG ATAGAAAATT TGCTGATATT GGTAATACCG
 AGTTAAATAC TAAAACCTTC TATCTTTAA ACCACTATAA CCATTATGCC

 4001 TAAAGAAAACA ATATATTGGT GGAGTTTATA AAATTAGTAG TTGGGCAGAT
 ATTTCTTGT TATATAACCA CCTCAAAATAT TTTAATCATC AACCCGTCTA

 4051 ATTACCAATG CTCATGGTGT CACTGGGAAT GGAGTGGTIG AAGGATTTAA
 TAATGGTAC GAGTACCCACA GTGACCCCTTA CCTCACCAAC TTCTAAATT

 4101 ACAGGGAGCT AAAGAAAACCA CCACCAACCA AGAGCCAAGA GGGTTATTGA
 TGTCCTCGA TTCTTTGGT GGTGGTTGGT TCTGGTTCT CCCAATAACT

 4151 TGTTAGCTGA ATTATCATCA GTGGATCAT TAGCATATGG AGAATATTCT
 ACAATCGACT TAATAGTAGT CACCTAGTA ATCGTATAACC TCTTATAAGA

 4201 CAAAAAACTG TTGAAATTGC TAAATCCGAT AAGGAATTG TTATTGGATT
 GTTTTTGAC AACTTTAACG ATTTAGGCTA TTCTTAAAC AATAACCTAA

 4251 TATTGCCAA CGTGATATGG GTGGCCAAGA AGAAGGATTG GATTGGCTTA
 ATAACGGTT GCACTATACC CACCGTTCT TCTTCCTAAA CTAACCGAAT

 4301 TTATGACACC TGGAGTGGT TTAGATGATA AAGGTGATGG ATTAGCACAA
 AATACTGTGG ACCTCAACCT AATCTACTAT TTCCACTACC TAATCCGTGTT

 4351 CAATATAGAA CTGTTGATGA AGTTGTTAGC ACTGGAACCTG ATATTATCAT
 GTTATATCTT GACAACCTACT TCAACAACTG TGACCTTGAC TATAATAGTA

 4401 TGTTGGTAGA GGATTGTTTG GTAAAGGAAG AGATCCAGAT ATTGAAGGTA
 ACAACCATCT CCTARCAAAC CATTTCCTTC TCTAGGTCTA TAACTCCAT

 4451 AAAGGTATAG AAATGCTGGT TGGAAATGCTT ATTTGAAAAA GACTGGCCAA
 TTTCATATC TTTACGACCA ACCTTACGAA TAAACTTTT CTGACGGTT

 4501 TTATAAATGT GAAGGGGGAG ATTTTCACTT TATTAGATTG GTATATATGT
 AATATTACA CTTCCCCCTC TAAAAGTGAATAATCTAA CATATATACA

 4551 AGAATAAATA AATAAAATAAG TAAATAATAAT AATTAATAAA GGGTGTAAAT
 TCTTATTAT TTATTATTC AATTATTTA TTAATTATTA CCCACCAATTA

 4601 TATTACTATT TACAATCAAA GGTGGTCCTT CTAGCTGTAAC TCCGGGAGC
 ATAATGATAA ATGTAGTTT CCACCAAGAA GATCGACATT AGGGCCGGTGC

 4651 GCAACCGAAC ATTCAATCAGT GTAAAAATGG AATCAATAAA GCCCTGGCA
 CGTTGCCTTG TAAGTAGTCA CATTCTTACCT TTAGTTATTT CGGGACGGT

 4701 GCGCCGAGGG TCACCTGAA TACGGTTTA ATGACCAAGCA CAGTCGTGAT
 CGCGCGTCCC AGTCGGACTT ATGCCAAAT TACTGGTCGT GTCAAGCACTA

4751 GGCAAGGTCA GAATAGCCCCA AGTCGGCCGA GGGCCCTGTA CAGTGAGGGA
CCGTTCCAGT CTTATCGGGT TCAGCCGGCT CCCCGGACAT GTCACTCCCT

4801 AGATCTGATA TTGACGAAGA GGACCAATG TAACGTTACA CTGAAGAAAA
TCTAGACTAT AACTGCTCT CCTGGTTAC ATTGCAATGT GACTTCTTT

4851 CACACAAATAA ACGGGAAGAA ACGGTGTAAA AGTGTGAAAA TAATTTTGA
GTGTGTATT TGCCCTCT TGCCACATTT TCACACTTTT ATAAAAAACT

4901 ATATCATTTC CCTTGGTTTA ATTCAAAACG AAACGTGTTT TTTTAGAGA
TATAGTAAAG GGAACCAAAT TAAGGTTTGC TTGACACAAA AAAAAATCTCT

EcoRI

ApalI

4951 ATGGGAATTTC TTATGGATG TCTAGATTGT TTGTTTACTC CAGACTCTGC
TACCTTAAG AATAACCTAC AGATCTAACAA AACAATGAG GTCTGACAGG

ApalI

5001 ACAAAAACGT TTGGATGGAT GATCAGAAGA TATTTTTAGG CTAGCTCTA
TGTGTTTGCA AACCTACCTA CTAGTCTCT ATAAAAATCC GAATCGAGAT

5051 AATAAAGAA ATGATGCTTG AAAAACAGA CAGAAATTGA GTTCAAAAAA
TTATATCTT TACTACGAAC TTTTGGTCT GTCTTTAATC CAAAGTTTT

5101 TTGGTAATGT GAGGTATTAG TCAACTAACC AAATAACAAT GCAAACCGGT
AACCATTACA CTCCATAATC AGTTGATTGG TTATATGTTA CGTTTGGCCA

5151 TGATACATTG CATTGAAA ATAATGAAAC TGGAAATTGG TGACCAAGC
ACTATGAAA GTAAAATCTT TATTACTTTG ACCTTAACCT ACTGGTCGIG

5201 ACAAACACAT AAAGTAATTA TGGGAATTAG AAGCGAACAT AGAGGAGTAC
TGTGTTGTA TTTCATTAAT ACCCTTAATC TTGCTTGTA TCTCTCATG

5251 TTGGCCACGA ACACAAATACA AGTGGGAACA CTATTTCTC CATIGTTTA
AACCCTGCT TGCTTATGT TCACCCCTGT GATAAAAGAG GTAACAAAAT

5301 GTTCTGTTT TTGTCAGCC TAGTTTTGTG CTATGTGAA AAAATATG
CAAGACAAA AAACAGTCGG ATCAAACAC GATACACATT TTGATAACG

HindIII

5351 CAAGAAAAAA AGCTTGTTT GTGGCCAGTG TCCGAAAAAA ATTTGGGGA
GTTCTTTTT TCGAACAAA CACCCGTAC AGCTTTTTT TAAACCCCT

5401 ATCTTCGGAT TAATTTATGT TTTCATTCCA TCGGGGAAAG TGGGGGGAA
TAGAAGCTA ATTAATACA AAAGTAAGGT AGCCCTTTC ACCCCCCCTT

5451 AAAATTTAA CGACTTCACA AAACCTTCCA AAAAATATAT GGACAAAGAT
TTTAAATT CGTCAAGTGT TTGGAAGGT TTTTATATA CCTGTTCTA

5501 GATTTGATT TCCCGACACC AAAATCATAA TTAATTATGA GAAAGTTAAA
CTAACATAAA AGGGCTGTGG TTTAGTATT AATTAATACT CTTCATTT

5551 TGTAACGTTA CAATTATGT TTATTTGAAG GTGAAAAGCG ATTTATGATT
ACATTGCAAT GTAAATACA AATAAACTTC CACTTTTCGC TAAATACTAA

5601 TTCCGAAAT GAAAATTTT TTAGGTTA TTTTTTGT CGGGCAAAGA
AAAGCTTTA CTTTAAAAA AATCCAAT AAAAANACCA GCCGTTCT

EP 0 982 401 A2

EcoRI

5651 AAAACTGAAC AAGGATTATT AAAATTTTG GTGTTGTTT GTGTCGGAG
 TTTGACTTG TTCTAATAA TTTAAAAAC CACAAACAAA CACAGACCTC

 EcoRI

 5701 AATTCAATTCC TCTCTCATCT TCACACAATG TTTAGACATC TGACACGGATT
 TTAAGTAAGG AGAGAGTAGA AGTGTGTTAC AATCTGTAG ACTGTGCTAA

 5751 CATGATAGTT CGGTTCCGG GGTTGGTGT TAGTTTGTGTT TTTCTTTTT
 GTACTATCAA GCCTAAAGGCC CCAACCAAA ATCAAAAGCA AAAAGAAAAA

 5801 TTTGGAAAG AATGTTTAG CTCATTGGTT TTCTTTCTTC ATTCAATAGT
 AAAACCTTTC TTACAAAATC GAGTAACCAA AAGAAAGAAG TAAGTATCA

 5851 TTTGAAAGAA TTGCCCCT TGTTATTACA ATCATATAAA ATAAACTT
 AAACTTCTT AAACGGGTGA ACAATAATGT TACTATATT TAATTGAAA

 5901 GATATAAAAT AGAGTTGAA AGTTCCAG ATCCTTTTG ATTTCTTGT
 CTATATTAA TCTCAACTT TCAAAGGGTC TAGGAAACAC TAAAGAAACA

 5951 AAATTTTTT TTCTCCACA TATACACACA TACAAACCGA TTTTATAAG
 TTAAAAAAA AAGAGGGTGT ATATGTGTT ATTTTGCT AAAAATATTC

PstI AvaI BamHI

6001 AAAGAGTTAT ACCCTGCAAGC TCGACCTCGA GGGATCCGGG CCCTCTAGAT
 TTCTCAATAA TGGAACGTCG AGCTGGAGCT CCCTAGGCC GGGAGATCTA

AvaI

6051 GGGCCCGCTA GCCTCGAGG GACTTTGCA CCAAAATAA TTATTTCC
 CGCCGGCGAT CGGAACTCC CTGAAAACGT GTTTTTATT AAATAAAAGG

 6101 AAAATAAAAT TTAAATAAAAT AAAATAACT CATAATTAA TAAAATTTC
 TTTTATTAA ATTTATTAA TTTTATTGA GTATTAATT ATTAAAAG

 6151 AAAATCTCT AGTGTCTT CATAATCCAGT ACATTAACCA TCAGTCACCT
 TTTAGAAGA TCACAGGAA GTATACGTCA TGTAAATCGGT AGTCAGTGAA

 6201 AAACAGCATC TGTGGTTGA AGAATGCTTG AAGCAATTGT CCAGTCCAG
 TTGTGGTAG AGCACAAC TCTACGAAAC TCTGTTAACAA GGTCAAGGGTC

 6251 AGGCACAGGC TAGGAGATCT TCAGTTTCGG AGGTAACCTG TAAGTCTGTT
 TCCGTGTCG ATCCCTCTAGA AGTCAAAGCC TCCATTGGAC ATTCAGACAA

 6301 AATGAAGTAA AACTTCTTAA SGATTTCCAC TCTGACTATG CTCCAGGCAC
 TTACTTCATT TTCAAGGAAT CCTAAAGGTG AGACTGATAC CAGGTCCGIG

 6351 AGTGACTGTA CTCTTGGCC TTCAGGTAAT GCAGAACCT CCCATAATAAT
 TCACTGACAT GAGGAACCGG AAGTCCATTCA CGTCTTAGGA GGGTATTATA

 6401 CTTTCAGGT GCAGACTGCT CATGAGTTT CCCCTGGTGA AATCTCTTT
 GAAAAGTCCA CGTCTGACGA TACTCAAA GGGGACCACT TTAGAAGAAA

 6451 CTCCAGTTT TCTTCAGGA CTGCTCTCAG ATGGTTTATC TGATGATAGA
 GAGGTCAAA AGAAGGTCT TACAGAAGTC TACCAAATAG ACTACTATCT

 6501 CATTAGCCAG GAGGTTCTCA ACAATAGTCT CATTCCAGCC AGTGCTAGAT
 GTAAATCGTC CTCCAAGAGT TTATCAGA GTAAGGTGG TCAAGGATCTA

EP 0 982 401 A2

6551 GAATCTTGTG TGAAAATAGC AAAGATGTTG TGGAGCATCT CATAGATGGT
CTTAGAACAG ACTTTTATCG TTCTACAAG ACCTCGTGA GTATCTACCA

PstI

6601 CAATGGGGG TCCTCTCTCT GGAACCTGCTG CAGCTGCTTA ATCTCTCTAG
GTTACGCCGC AGGAGGAAGA CCTTGACGAC GTGAGGAAT TAGAGGAGTC

6651 GGATGTCAAA GTTCATCTG TCCCTGAGGC AGTATTCAG CCTCCCATTC
CCTACAGTTT CAAGTAGGAC AGGAACCTCG TCATAAGTTG GGAGGGTAAG

6701 AATTGCCACA GGAGCTCTG AACTGAAAAA TTGCTGCTTC TTGTAGGAA
TTAACGGTGT CCTCGAAGAC TGTGACTTT AACGACGAAG AACATCTT

6751 TCCAAGCAAG TTGTAGCTCA TGGAAAGAGC TGTAGTGGAG AAGCACACA
AGGTGGTTC AACATCGAGT ACCTTTCTG ACATCACCTC TTCTGTGTTG

AvaI

6801 GGAGAGCAAT TTGGAGGAGA CACTTGTGG TCATGTTCCCT CGAGGCCCTT
CCTCTGTTA AACCTCTCT GTGACAACACC AGTACAAGGA GCTCCGGAAA

BamHI

6851 TTGGCCAGCT GCGGCGCTGCT GCGCGACGGC GAGCTGCTCA CCACCCAGGA
AACCGGTGCA CGCGCGACGA CGCGCTGCCG CTCGACGAGT GGTGGGTCT

BamHI

6901 TCCGTCCCCC TTTCCTTGTG TCGATATCAT GAAATTAGTT ATGTCACGCT
AGGCAGGGGG AAAAGAAC AGCTATAGTA CATTAAATCAA TACAGTGGAA

6951 TACATTCAAG CCCTCCCCC ACATCCGCTC TAACCGAAAA GGAAGGAGGT
ATGTAAGTGC GGGAGGGGG TGTAAGGAG ATTGGCTTT CCTTCCCTCAA

7001 AGACAACCTG AAGTCTAGGT CCCTATTTAT TTTTTATAG TTATGTTAGT
TCTGTTGGAC TTCAGATCCA GGGATAAATA AAAAAATATC AATACAATCA

7051 ATTAAGAACG TTATTTATAT TTCAAAATTT TCTTTTTTT CTGTACAGAC
TAATTCTTGC AATAAAATATA AAGTTAAAAA AGAAAAAAA GACATGTCG

7101 GCGGTGACGC ATGTAACATT ATACTGAAAA CCTGCTTGA GAAGGTTTG
CCCACATGGC TACATTGAA TATGACTTTT GGAAACGAACT CCTCCAAAAC

HindIII

7151 GGACCGCTCGA AGGCTTTAAT TTGCA
CCTGCGAGCT TCCGAAATTA AACGT

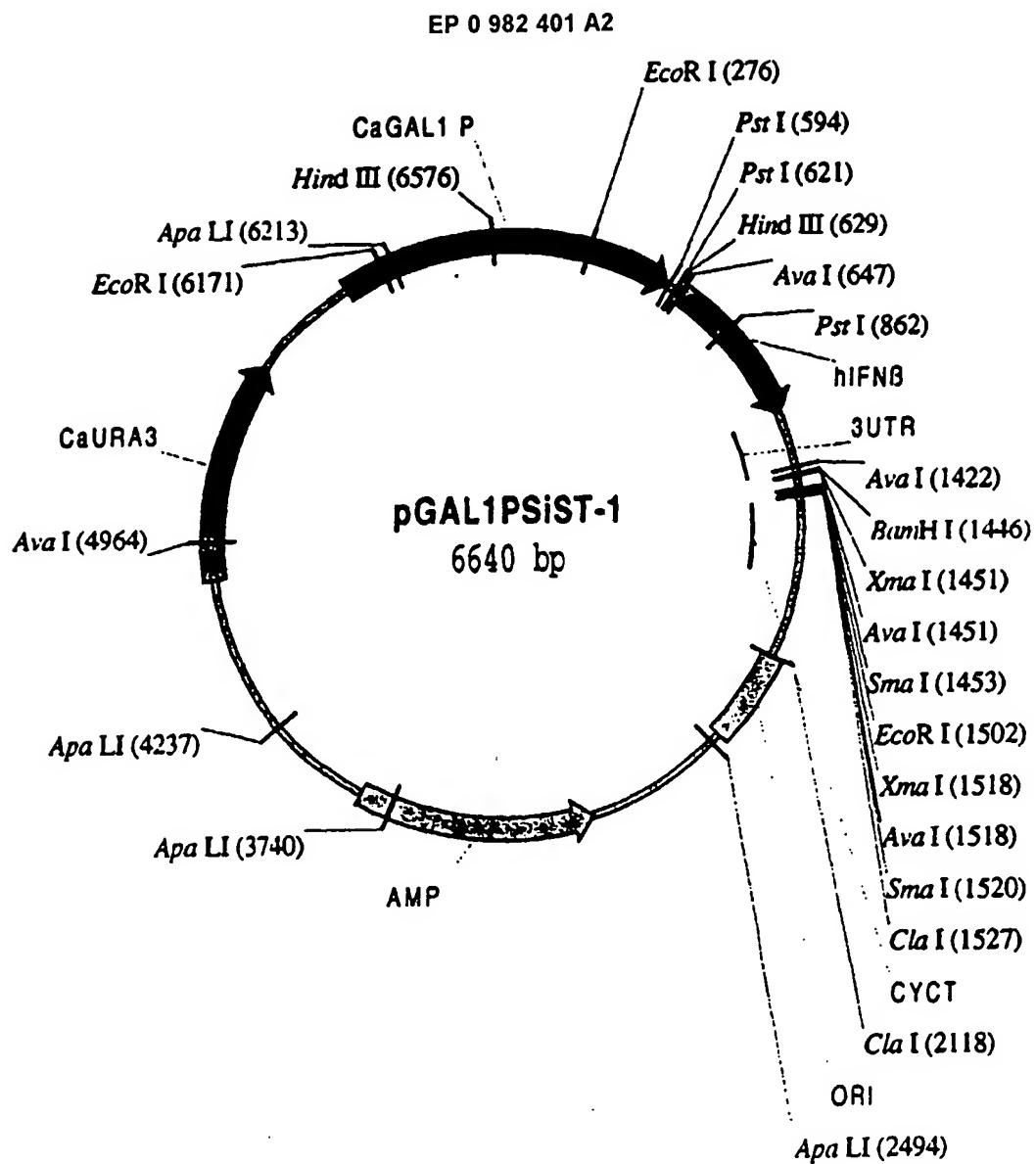


Fig 3

EP 0 982 401 A2

1 TTCCCATCGGG GAAAGTGGGG GGGAAAAAAAT TTTAAGCAGT TCACAAAACC
 AAGGTAGCCC CTTCACCCC CCCTTTTTA AAATTCGTCA AGTGTGTTGG

 51 TTCCAAAAAA TATATGGACA AAGATGATTC TATTTTCCCG ACACCAAAAT
 AAGGTTTTT ATATACTGT TTCTACTAAC ATAAAAGGGC TGTTGTTTTA

 101 CATAATTAAT TATGAGAAAG TAAATGTAA CGTACAAATT TATGTTTATT
 GTATTAATTA ATACTCTTC AATTACATT GCAATGTAA ATACAAATAA

 151 TGAAGGTGAA AACCGATTAA TGATTTTCC GAAATGAAA TTGTTTTAG
 ACTTCCACTT TTGCTAAAT ACTAAAAGG CTTTACTTT AAAAAGAAATC

 201 GTTTATTTT TTGTCGGGC AAAGAAAAAC TGAACAAGGA TTATTAATAA
 CAAATAAAA AAACAGCCCC TTCTTTTG ACTTGTCTT AATAATTTA

EcoRI

251 TTTGGGTGTT TGTGTTGTC TGGAGAATTG ATTCCTCTC CATCTTCACA
 AAAACCACAA ACAACACAG ACCTCTTAAG TAAGGAGAGA GTAGAAGTGT

 301 CAATGTTAG ACATCTGACA CGATTCTGAG TAGTTGCGTT TCCGGGGTTG
 GTTACAAATC TGAGACTGT GCTAAGTACT ATCAAGCCAA AGGCCCCAAC

 351 GTGTTAGT TTGTTTTTC TTTTTTTTG GAAAGAATGT TTAGCTCAT
 CACAAATCAA AAGCAAAAG AAAAAGAAAC CTTCTTACA AAATCGAGTA

 401 TGGTTTCTT CTTCATTCA ATAGTTTGA AAGAATTGCA CCACCTGTTA
 ACCAAAAGAA AGAAGTAAGT TATCAAACAT TTCTAAACG GGTGAACAAT

 451 TTACAATCAT ATAAAATTAA ACTTTGATAT AAAATAGAGT TTGAAAGTTT
 AATGTTAGT TATTTTAATT TGAAACTATA TTTTATCTCA AACTTTCAA

 501 CCCAGATCCT TTTGATTTC TTGTAATT TTTTTCTC CCACATATAC
 GGGCTAGGA AAAACTAAAG AAACATTAA AAAAAGAG GGTGTATATG

PstI

551 ACACATACAA ACCGATTTT ATAAGAAAAGA GTTATACCC GCAAGCTCGAC
 TGCTATGTT TGGCTAAAAA TATTCTCTC CAATATGGGA CGTCGAGCTG

PstI HindIII

AvaI

601 CTCGACTGTT TAAACCTGCA GGCATGCAAG CTGGCCAAA AAGGCTCGA
 GAGCTGACAA ATTTGGACGT CGCTACGTT GAAACGGTTT TTGCGGAGCT

 AvaI

 651 GGAACATGAC CAACAAGTGT CTCCCTCAA TTGCTCTCCT GTTGTGCTTC
 CCTGTACTG GTTGTCTACA GAGGAGTTT AACGAGAGGA CAACACGAAG

 701 TCCACTACAG CTCTTCCAT GAGCTACAC TTGCTGGAT TCCACAAAG
 AGGTGATGTC GAGAAAGGT ATCGATGTTG AACGAACCTA AGGATGTTTC

 751 AAGCAGCAAT TTGACTGTC AGAACCTCCT GTGCCAATG AATGGAGGC
 TTGTCGTTA AAAGTCACAG TCTTCGAGGA CACCGTTAAC TTACCCCTCG

 801 TTGAATACTG CCTCAAGGAC AGGATGAAC TGTACATCCC TGAGGAGAT
 AACTTATGAC GGAGTTCTG CCTACTTGA AACTGTAGGG ACTCCTCTAA

Fig 4

EP 0 982 401 A2

PstI

851 AAGCAGCTGC ACCAGTTCCA GAAGGAGGAC CCCGCATTGA CCATCTATGA
TTCGTCGACG TCGTCAAGGT CTTCTCTG CGCGTAACT GGTAGATACT
.....
901 GATGCTCCAG AACATCTTG CTATTTTCAG ACAAGATTCA TCTAGCACTG
CTACGAGGTC TTGTAGAAC GATAAAAGTC TGTCTAAAGT AGATCGTGAC
.....
951 GCTGGAATGA GACTATTGTT GAGAACCTCC TGGCTAATGT CTATCATCAG
CGACCTACT CTGATAACAA CTCTTGAGG ACCGATTACA GATAGTAGTC
.....
1001 ATAAACCACAT TGAAGACAGT CCTGGAAGAA AAACCTGGAGA AAGAAGATTT
TATTTGGTAG ACTTCTGTCA GGACCTCTT TTGACCTCT TTCTCTAAA
.....
1051 CACCAGGGAA AAACATCATGA GCAGTCTGCA CCTGAAAAGA TATTTATGGG
GTGGTCCCCCT TTTGAGTACT CGTCAGACGT GGACTTTCT ATAATACCCCT
.....
1101 GGATTCTGCA TTACCTGAAG GCCAAGGAGT ACAGTCACTG TGCCCTGGACC
CTTAAGACGT AATGGACTTC CGGTTCTCA TGTCACTGAC ACGGACCTGG
.....
1151 ATAGTCAGAG TGGAAATCCT AAGGAACATT TACTTCATTA ACAGACATTAC
TATCAGTCACCTT ACCTTTAGGA TTCTTGAAA ATGAAGTAAT TGTCTGAATG
.....
1201 AGGTTACCTC CGAAACTGAA GATCTCTAG CCTTGCCCTC TGGGACTGGA
TCCAATGGAG GCTTTGACTT CTAGAGGATC GGACACGGAG ACCCTGACCT
.....
1251 CAATTGCTTC AAGCATTCTT CAACCAAGCAG ATGCTGTTA AGTGACTGAT
GTTAACGAAG TTGTAAGAA GTTGGTGTGTC TACGACAAAT TCACTGACTA
.....
1301 GGCTAATGTA CTGGATATGA AAGGACACTA GAAGATTTG AAATTTTAT
CCGATTACAT GACGTATACT TTCTGTGAT TTCTAAAAC TTAAAAATA
.....
1351 TAAATTATGA GTTATTTTA TTTATTTAA TTTATTTG GAAAATAAT
ATTTAATACT CAATAAAAAT AAATAAATTT AAATAAAAAC CTTTATTTA
.....

XmaI

SmaI

BamHI

AvaI

AvaI

1401 TATTTTTGGT GCAAAAGTCC CTCGAGGCCT AGCGCCGCC TAGAGGATCC
ATAAAAACCA CGTTTCAGG GAGCTCCGGA TCGCCGGCGG ATCTCTAGG
.....

XmaI

SmaI

AvaI

1451 CGGGGCGCTA GGCAGCCGCT AGGCCTTTT GGCCAGCTC GAATTCGAG
GGCCCGCGAT CGCCGGCGA TCCGAAAAAA CCGGTTCGAG CTAAAGCTC
.....

XmaI

SmaI

EcoRI

AvaI Clai

1501 GAATTGGAGC TCGGTACCCG GGGGATCGAT CCGTCCCCCT TTTCCTTGT
CTTAAGCTCG AGCCATGGGC CCCCTAGCTA GGCAGGGGA AAAGGAAACA
.....

EP 0 982 401 A2

1551 CGATATCATG TAATTAGTTA TGTCAACGTT ACATTCACCC CCTCCCCCA
GCTATAGTAC ATTAATCAAT ACAGTGCAGA TGTAAAGTGCAG GGAGGGGGGT
.....

1601 CATCCGCTCT AACCGAAAAG GAAGGAGTTA GACAACCTGA AGTCTAGGTC
GTAGGGAGA TTGGCTTTTC CTTCCTCAAT CTGTTGACT TCAGATCCAG
.....

1651 CCTATTATT TTTTATAGT TATGTTAGTA TPAAGAACGT TATTATATT
GGATAAATAA AAAATATCA ATACAATCAT AATTCTTGCA ATAATAATAA
.....

1701 TCAAATTATT CTTTTTTTC TGTACAGACG CGTGTACGCA TGTAAACATTA
AGTTAAAAA GAAAAAAAG ACATGTCAGC GCACATCCGT ACATTGTAAT
.....

1751 TACTGAAAAC CTGGCTTGAG AAGGTTTGG GACGGCTCGAA GGCTTTAATT
ATGACTTTG GAACGAACTC TTCCAAAACC CTGGGAGCTT CGGAAATTAA
.....

1801 TGCAAGCTAG CTGGCGTAA TCATGGTCAT AGCTGTTTCC TGTGTGAATT
ACGTTCGATC GAACCGCATT AGTACCCAGTA TCGACAAAGG ACACACTTTA
.....

1851 TGTTATCCG TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG
ACAAATAGGG AGTGTAAAG TGTGTTGTAT GCTGGCCCTT CGTATTTCAC
.....

1901 TAAAGCCTGG GGTGCTTAAT GAGTGAGCTA ACTCACATTA ATTGGGTTGC
ATTTCGGACC CCACGAGTTA CTCACTCGAT TGAGTGTAAAT TAAACGCAACG
.....

1951 GCTCACTGCC CGCTTTCCAG TCGGGAAACC TGTGCTGCCA GAGATCTCTG
CGAGTGACGG CGGAAAGTC AGCCCTTGG ACAGCACGGT CTCTAGAGAC
.....

2001 CATTAAATGAA TCGGCCAACG CGCGGGGAGA GGGGGTTTGC GTATTGGGG
GTAATTACTT AGCCGGTTGC CGCCCTCTC CCCGAAACG CATAACCCCG
.....

2051 CTCTTCGGT CTCTCGCTCA CTGACTCGCT GCGCTGGTC GTTGGCTOC
GAGAAGGGGA AGGAGGGAGT GACTGAGCGA CGCGAGCCAG CAAGCCGACG
.....

ClaI

2101 GGCAGGGGT ATCAGATCGA TCTCACTCAA AGGGGGTAAT ACGGTTATCC
CCGCTCCCA TAGTCTAGCT AGAGTGAGTT TCCGCCATTA TGCCAATAGG
.....

2151 ACAGAACATCG GGGATAACCG AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA
TGTCTTAGTC CCCTATGGG TCCCTTCTTG TACACTCGTT TTCCGGTCGT
.....

2201 AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGGTTT TTCCATAGGC
TTCCGGTCC TTGGCAATT TCCGGGCAAA CGACCCAAA AAGGTATCCG
.....

2251 TCCGCCCTCC TGACGAGCAT CACAAAAATC GACCGCTCAAG TCAGAGGTGG
AGGCGGGGGG ACTGCTCGTA GTGTTTTAG CTGGAGTTTC AGTCTCCACC
.....

2301 CGAAACCCGA CAGGACTATA AAGATACCAAG CGTTTCCCC CTGGAAAGCTC
GCTTTGGCT GTCTGTATAT TTCTATGGTC CGCAAAGGGG GACCTTCGAG
.....

2351 CCTCGTGCCTC TCTCTCTTC CGACCCCTGCC GCTTACCGGA TACCTGTCCG
GGAGCACCGG AGAGGACAAG GCTGGGACGG CGAATGGCCT ATGGACAGGC
.....

2401 CCTTTCTCCC TTGGGAAGC GTGGCGCTTT CTCATAGCTC ACCTGTAGG
GGAAAGAGGG AAGCCCTCG CACCGGAAA GAGTATCGAG TGGGACATCC
.....

ApaLI

2451 TATCTCAGTT CGGTGTAGGT CGTTGGCTCC AAGCTGGGT GTGTGCACCA
ATAGAGTCAA CCCACATCCA GCAAGCGAGG TTGGACCCGA CACACGTGCT
.....

2501 ACCCCCGT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG
 TGGGGGCAA GTCCGGCTCG CGACGGGAA TAGGCCATTG ATAGCAGAAC
 .
 2551 AGTCCAACCC GGTAAACAC GACTTATCGC CACTGGCAGC AGCCACTGGT
 TCAGGTGGG CCATTCTGTG CTGAATACGG GTGACCGCTCG TCGGTGACCA
 .
 2601 AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA
 TTGTCTTAAT CGTCTCGTC CATAACATCGG CCAACGATGTC TCAAGAACTT
 .
 2651 GTGGTGCCCT AACTACGGCT AACTAGAAAG GACAGTATTT GGATCTCGCG
 CACCACCGA TTGATGCCGA TGTGATCTTC CTGTCATAAA CCATAGACGC
 .
 2701 CTCTGCTGAA GCCAGTTACC TTGGAAAAAA GAGTTGGTAG CTCTTGATCC
 GAGACGACTT CGCTCAATGG AAGCTTTTT CTCACCCATC GAGAACTAGG
 .
 2751 GGCAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTGT GCAACCGAGCA
 CGGTTGTTT GGTGGCGACC ATGCCACCA AAAAAAGAAA CGTTCGTCGT
 .
 2801 GATTACGGC AGAAAAAAAG GATCTCAAGA AGATCTTTTG ATCTTTCTA
 CTAATGCCGG TCTTTTTTC CTAGAGTTCT TCTAGGAAAC TAGAAAAGAT
 .
 2851 CGGGGCTCTGA CGCTCAGTGG AACGAAAAGT CACGTTAAGG GATTTGGTC
 GCCCCAGACT GCGAGTCACC TTGCTTTGTA GTGCATTTC CTAACCCAG
 .
 2901 ATGAGATTAT CAAAAGGAT CTTCACCTAG ATCCTTTAA ATTAAAATG
 TACTCTAATA GTTTTCTTA GAAGTGGATC TAGGAAAATT TAATTTTAC
 .
 2951 AAGTTTAA TCAATCTAA GTATATATGA GTAAACTTGG TCTGACAGTT
 TTCAAAATT AGTTAGATT CATATATACT CATTGAACC AGACTGTCAA
 .
 3001 ACCAATGCTT AATCACTGAG GCACCTATCT CAGCGATCTG TCTATTTGCT
 TGGTTACGAA TTAGTCACTC CGTGGATAGA GTGGCTAGAC AGATAAAGCA
 .
 3051 TCATCCATAG TTGCTGACT CCCCGTCGTG TAGATAACTA CGATAACGGGA
 AGTAGGTATC AACGGACTGA GGGCAGCAC ATCTATGAT GCTATGCCCT
 .
 3101 GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATAACCGGA GACCCACGCT
 CCCGAATGGT AGACGGGGT CACGACGTTA CTATGGCGCT CTGGGTGCGA
 .
 3151 CACCGGCTCC AGATTATCA GCAATAAACC AGCCACCCGG AAGGGCCGAG
 GTGGCCGAGG TCTAAATAGT CGTTATTTGG TCGGTGGCC TTCCCGGCTC
 .
 3201 CGCAGAAGTG GTCTGCAAC TTTATCCGGC TCCATCCAGT CTATTAATG
 GCGTCTCAC CAGGACGTTG AAATAGGGGG AGTAGGTCA GATAATTAAC
 .
 3251 TTGGGGGAA GCTAGAGTAA GTAGTTGGCC AGTTAATAGT TTGCGCAACG
 AACGGCCCTT CGATCTCATT CATCAACGGG TCAATTATCA AACGGTTGC
 .
 3301 TTGTTGCCAT TGCTACAGGC ATCGTGGGT CACGCTCGTC GTTTGGTATG
 AACAAACGTA ACAGATGTCGG TAGCACCACCA GTGGGAGCAG CAAACCATAC
 .
 3351 GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGGGACTTA CATGATCCCC
 CGAAGTAAGT CGAGGCCAAG GTTTGCTAGT TCGGTCAAT GTACTAGGGG
 .
 3401 CATGTTGTC AAAAAGGGG TTAGCTCTT CGGTCTCCG ATCGTTGTCA
 GTACAAACAGG TTTTTGCCC AATCGAGGAA GCGAGGAGGC TAGCAACAGT
 .
 3451 GAAGTAAGTT GCGCCGACTG TTATCACTCA TGGTTATGGC AGCACTGCAT
 CTTCAATTCAA CGGGGTCAAC AATAGTGAGT ACCAATACCG TCGTGACGTA

EP 0 982 401 A2

3501 AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTCTG TGACTGGTGA
 TTAAGAGAAT GACAGTACGG TAGGCATTCT AGGAAAAGAC ACTGACCACT

 3551 GTACTCAACC AAGTCATTCT GAGAATAGTG TATCGGGCGA CGGAGTGTCT
 CATGAGTTGG TTCAGTAAGA CTCTTATCAC ATACGCCGCT GGTCAACGA

 3601 CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA
 GAACGGCCCG CAGTTATGCC CTATTATGGC GGGGTGTATC GTCTTGAAT

 3651 AAAGTGCTCA TCATTGGAAA ACGGTCTTCG GGGCGAAAAC TCTCAAGGAT
 TTTCACGAGT AGTAACCTTT TGCAAGAAGC CCCCTTTTG AGAGTTCTA

ApaLI

3701 CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTGTC GCACCCAAC
 GAATGCCGAC AACTCTAGGT CAAGCTACAT TGGGTGAGCA CGTGGGTGTA

 3751 GATCTTCAGC ATCTTTACT TTCAACCAGCG TTCTGGGTG AGCAAAAACA
 CTAGAAGTGTG TAGAAAATGA AAGTGGTCCC AAAGACCCAC TGTTTTGTG

 3801 GGAAGGCAAATGGCCAAA AAAGGGAATA AGGGCGACAC GGAAATGTTG
 CCTTCCGTTT TACGGCGTTT TTCCCTTAT TCCCGCTGTG CCTTTACAAC

 3851 AATACTCATA CTCTTCCATT TTCAATATTA TTGAGGATT TATCAGGTT
 TTATGAGTAT GAGAAGGAAA AAGTTATAAT AACTTCGTAA ATAGTCCCAA

 3901 ATTTGCTCAT GACGGGATAC ATATTGAAAT GTATTAGAA AAATAAACAA
 TAACAGAGTA CTGGCTATG TATAAACTTA CATAAATCTT TTATTTGTT

 3951 ATAGGGGTTTC CGCCGACATT TCCCGAAAAA GTGCCACCTG ACGTCTAAGA
 TATCCCAGG GCGCGTGTAA AGGGGCTTTT CACGGTGGAC TGCAGATTCT

 4001 AACCATTATT ATCATGACAT TAACCTATAA AAATAGGGT ATCACGAGGC
 TTGGTAATAA TAGTACTGTA ATTGGATATT TTATCCGCA TAGTGGTCCG

 4051 CCTTTCGTCT CGCGCGTTTC GTGATGACG GTGAAAACCT CTGACACATG
 GGAAAGCAGA CGCGCAGAAG CCACTACTGC CACTTTGGA GACTGTGTAC

 4101 CAGCTCCCG AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCG
 GTCGAGGGCC TCTGCCAGTG TCGAACAGAC ATTGCGCTAC GGCCTCGTC

 4151 ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGCGGGTGT CGGGGCTGGC
 TGTTGGGCA GTCCCCGCA GTCCCCACAA ACCCCCCACA GCGGGGACCG

ApaLI

4201 TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATCGAC
 AATTGATAGG CCGTAGTCG GTCTAACATG ACTTCACGT GGTATACTG

 4251 GCTCTCCCTT ATCGGACTCC TGCATTAGGA AGCAGCCCAG TAGTAGGTG
 CGAGAGGGAA TACGCTGAGG ACGTAATCTT TCGTGGGTGTC ATCATCCAAC

 4301 AGGCCGTTGA GCACCGCCGC CGCAAGGAAT GTGCCATGCA AGGGAGATGGC
 TCCCGCAACT CGTGGCGGGCG GGGTCTTAA CCACGTACGT TCTCTACCG

 4351 GCCCCACAGT CCCCCGGCCA CGGGGCCCTGC CACCATACCC AGGCCGAAAC
 CGGGTTGTCA GGGGGCCCGT GCCCCGGACG GTGGTATGGG TCGGGCTTTC

 4401 AAGCACTAAT AGGAATTGAT TTGGATGGTA TAAACGGAAA CAAAAAAAAG
 TTCTGTGATTA TCCTTAACCA AACCTACCAT ATTGCTTTT GTTTTTTTC

EP 0 982 401 A2

4451 AGCTGGTACT ACTTTCTTTA AATTATTTT ATTATTTGAT TTATTTAAT
 TCGACCATGA TGAAAGAAT TTTAATAAA TAATAAACA AATAAAATTA

 4501 AGTATATATT ATATTTGAA CGTAGATTAT TTGTTGAAA GTTGTGTAG
 TCATATATA TATAAACTT GCATCTAATA AAACACTTT CAACGACATC

 4551 TGCCATGAT TCGTAACACT AATTCTGTAT TAGTCATTCC TCTGTTGA
 ACGGTAACCA AGCATTTGTA TTAAGACATA ATCGTAAGG AGAACAAACT

 4601 TAGTATCCAA AAAAACGGCT ATTTTTTGC AATCTTATT CCTGCATATT
 ATCATAGTT TTTTGCGA TAAAAAAACG TTAGAATAAA GGACGTATAA

 4651 ATACAGATAA CATAATGAA GAAAAAATCT TTTTTTTGT TCTTCATGA
 TATGTCATT GTATTACTTT CTTTTTAGA AAAAAAAACA AGAAGTACT

 4701 TGATTTCAAC CATTCTTTA AACATTGATC AATTCTGAG CAACAAACCC
 ACTAAAGTG GTAAAGAAT TTGTAACTAG TTAAGGACTC GTTGTGGGG

 4751 ATACACACTG GTTATATAC CGCCCCTTT ACACCTGAG AAAGAAATAG
 TATGTGTGAC CAATATATG GCGGGAAAAA TGTCACCTTC TTTCCTTATC

 4801 AAATAGAAAAT AGCAAACAAA AGATATGACA GTCAACACTA AGACCTATAG
 TTTATCTTTA TCGTTGTT TCTATACTGT CAGTTGTGAT TCTGGATATC

 4851 TGAGAGACCA GAAACTCATG CCTCACCACT AGCACAGCGA TTATTTGAT
 ACTCTCTCGT CTTTGACTAC GGAGTGGTCA TCGTGTGCGT AATAAAGCTA

 4901 TAATGAACT GAAGAAAACC AATTATGTC CATCAATTGA CGTTGATACC
 ATTACCTTGA CTTCCTTTGG TTAATACAC GTAGTAACT CCAACTATAC

AvaI

4951 ACTAAGGAGT TCCTCGAGTT AATTGATAAA TTAGGTCTT ATGTATGCTT
 TGATTCCTCA AGGAGCTAA TTAACTATTT AATCCAGGA TACATACGAA

 5001 AATCAAGACT CATATTGATA TAATCAATGA TTTTCCTAT GAATCCACTA
 TTAGTTCTGA GTATAACTAT ATTAGTTACT AAAAGGATA CTTAGGTGAT

 5051 TTGAAACCTT ATTAGAACTT TCACGAAAC ATCAATTAT GATTTTGAA
 AACTGGTAA TAATCTGAA AGTCATTTG TAGTTAAATA CTAAAAACTT

 5101 GATAGAAAAT TTGCTGATAT TGGTAATACC GTAAAGAAAC AATATATTGG
 CTATCTTTA AACGACTATA ACCATTATGG CATTCTTTG TTATATAACC

 5151 TGGAGTTAT AAAATTAGTA GTTGGGCAGA TATTACCAAT GCTCATGGTG
 ACCTCAAATA TTTTAATCAT CAACCGCT ATAATGGTTA CGAGTACAC

 5201 TCACTGGAA TGGAGTGGTT GAAGGATTA AACAGGGAGC TAAAGAAAACC
 AGTGACCCCTT ACCTCACCAA CTTCCTAATT TTGTCCTCG ATTCTTTGG

 5251 ACCACCAACC AAGAGCCAAG AGGGTTATTG ATGTTAGCTG AATTATCATC
 TGGTGGTGGT TCTCGGTT TCCCAATAAC TACAATCGAC TTAATAGTAG

 5301 AGTGGGATCA TTAGCATATG GAGAATATTG TCAAAAAACT GTGAAATTG
 TCACCCCTACT AATCGTATAC CTCTTATAAG AGTTTTTGA CAACTTTAAC

 5351 CTAAATCCGA TAAGGAATTG GTTATGGAT TTATGCCCCA ACGTGATATG
 GATTTAGGCT ATTCCCTAAA CAATAACCTA AATAACGGGT TCCACTATAC

5401 GGTGGCCAAG AAGAAGGATT TGATGGCTT ATTATGACAC CTGGAGTTGG
 CCACCGGTTTC TTCTTCTAA ACTAACCGAA TAATACTGTG GACCTCAACC

 5451 ATTAGATGAT AAACGTGATG GATTAGGACA ACAATATAGA ACTGTTGATG
 TAATCTACTA TTTCACATAC CTAATCCCTGT TGTATATCT TGACAACATAC

 5501 AAGTTGGTAG CACTGGAAC GATATTATCA TTGTTGGTAG AGGATTGTTT
 TTCAACAATC GTGACCTTGA CTATAATAGT AACAAACCATC TCCTAACAAA

 5551 GGTAAAGGAA GAGATCCAGA TATTGAAGGT AAAAGGTATA GAAATGCTGG
 CCATTTCTT CTCTAGGTCT ATAACCTCCA TTTCCATAT CTTTACGACC

 5601 TTGGAATGCT TATTGAAAAA AGACTGGCCA ATTATAATG TGAAGGGGG
 AACCTTACCA ATAACCTTT TCTGACCGGT TAATATTAC ACTTCCCCCT

 5651 GATTTTCACT TTATTAGATT TGTATATATG TAGAATAAAAT AAAATAATAA
 CTAAAAGTGA ATAATCTAA ACATATATAC ATCTTATTTA TTATTTATT

 5701 GTTAAATAAA TAATTAATAA AGGGTGGTAA TTATTAACTAT TTACAATCAA
 CAATTATTT ATTAAATTTAT TCCCACCATT AATAATGATA AATGTTAGTT

 5751 AGGTGGCTT TCTAGCTGTA ATCCGGGAG CGCAACGGAA CATTTCATCAG
 TCCACCAAGGA AGATCGACAT TAGGGCCGTC GCGTTGCCCT GTAAGTAGTC

 5801 TGAAAAATG GAATCAATAA AGCCCTGGCG TCATGAGCCC GAAGTGGCGA
 ACATTTTAC CTAGTTATT TGGGACCGG AGTACTCGGG CTTCACCGCT

 5851 GCGCGATCTT CCCCATCGGT GATGTCGGCG ATATAGCGC CAGCAACCGC
 CGGGCTAGAA GGGTAGCCA CTACAGCGC TATATCCCGC GTCTGGCG

 5901 ACCTGTGGCG CGCGAGCGCG CAGGGTCAGC CTGAATACGC GTTTAATGAC
 TGGACACCGC GCGGTGCGC GTCCCACTCG GACTTATGCG CAAATTACTG

 5951 CAGCACACTC GTGATGGCAA GGTCAAGATA GCCCAAGTCG GCGGAGGGGC
 GTCGTGTAG CACTACCGTT CCAGTCTTAT CGGGTTTCAGC CGGCTCCCCG

 6001 CTGTACAGTG AGGGAAAGATC TGATATTGAC GAAGAGGAAC CAATGTAACG
 GACATGTAC TCCCTCTAG ACTATAACTG CTTCTCTTG GTTACATTG

 6051 TTACACTGAA GAAAACACAC AATAAACGGG AAGAAACGGT GTAAAAAGTGT
 AATGTGACTT CTTTTGTGTG TTATTTGCCC TTCTTGCCTA CATTTCACAA

 6101 GAAAATAATT TTGAATATC ATTTCCCTTG GTTAAATTCC AAACGAAACG
 CTTTTATTAAG AACTTATAG TAAAGGAAAC CAAATAAGG TTGCTTTGCC

EcoRI

6151 TGTTTTTTT AGAGAATGGG AATTCTTATT GGATGTCTAG ATTGTTTGT
 AAAAAAAA TCTCTTACCC TTAAGAATAA CCTACAGATC TAACAAACAA

ApalI

6201 TACTCCAGAC TGTGCACAAA AACGTTGGGA TGGATGATCA GAAGATATT
 ATGAGGTCTG ACACGTGTTT TTGCAAAACCT ACCTACTAGT CTTCTATAAA

 6251 TTAGGCTTAG CTCTAAATAT AAGAAATGAT GCTGAAAAA CCAGACAGAA
 AATCCGAATC GAGATTATA TTCTTACTA CGAACTTTT GGTCTGTCTT

 6301 ATTGAGTTTC AAAAATTGGT AATGTGAGGT ATTAGTCAAC TAACCAATA
 TAACTCAAAG TTTTAACCA TTACACTCCA TAATCAGTTG ATTGGTTAT

EP 0 982 401 A2

6351 ACAATGCAAA CGGGTTGATA CATTTCATTT TGAAAATATA GAAACTGGAA
TGTTACGTTT GGCCAACTAT GTAAAGTAAA ACTTTTATAA CTTTGACCTT

6401 TTGGATGACC AGCACACAAA CACATAAACT AATTATGGGA ATTAGAGCG
AACCTACTGG TCGTGTGTTT GTGTATTTCG TTAATACCCCT TAATCTTCGG

6451 AACATAGAGG AGTACTTGGC CACGAAACAGA ATACAAGTGG GAAACACTATT
TTGTATCTCC TCATGAACCG GTGCTTGCT TAATGTCACC CTGCTGATAA

6501 TTCTCCATTG TTTAGTTCT GTTTTTTGT CACCCCTAGTT TTGTGCTATG
AAGAGGTAAAC AAAATCAAGA CAAAAAAACA GTGGGATCAA AACACGATAC

HindIII

6551 TGTAAGGAAAT ATTGCCAAGA AAAAAGGCTT GTTTTGTTGGC CAGTGTCGGA
ACATTTTTA TAACCGTTCT TTTTTTCGAA CAAACACCG GTCACAGGCT

6601 AAAAAAATTTT GGGGAATCTT CGGATTAATT TATGTTTCA
TTTTTTAAAA CCCCTTAGAA GCCTTAATTAA ATACAAAAGT

Sequences with unknown function, *C. albicans* sequence NOT present in the public domain
(ALCES/EMBL)

>328c2 1803bp in-house: 1123-1803 public: 1-436/468-1021 PathoSeq:
437-467/1022-1122

ATGTCTATTACAGTTACATTCCGAAATCTCCATCTACGAAAAACGTGCACCG
GCATTGGAATTGAGTTGGAGTTAG
TCAMCAAGSCAGTAGCGATGGTGCTATAGAGAAAAGCGGCATTGGCAGTTCC
GTGTTAGCGTTGACAACCAAGACTWT
GTATTKATAAGAGAYCWTGCCAAGTACTGGGGCTACCCCTCATCGTATCAATT
GATTGTCAAGTTGGTCAAATGTGCTAA
CATTGAAAAGTCGCAAATCTAAAGACCGATAAGGATTGAATAGAGAGTTGT
TTGAGTTGGATTGATTGAAGAAGCAG
ATACAAAGATTGATCTTTTATATTGTTACCCCTGGTCTATTCAAGAATAGA
AAATAAGAAGGTTTTATGTTCTG
CGTGAACCAGAACAGCCAAAGGTGTCGAAAGCMCCAACACAAGAGAAACCAG
CAAGTGTGGTTGCTGCAGAAGAAGATGA
CGATAATCTAGATGATGATGAGGAGGACGAAGTGGATGAAGACATGGATGAA
GATAATGATAATAGTGGGAATTGTCTA
AAGGATACAAGCACATGCACAAGGACATCCAAAGTATATAAATGACCGATAG
GGTTACTATTGGACAAGTGTTCATCAA
TACGGACTTGACCCCTCGACACCATTAAACCCATTCACTTTCAATAGTATCAAC
TCAATGTCGAAGCTAAACTATTACAA
GAATTGGAGTTCAAGGTTACCGATTCTCCCAACAGCAAGTTATCTTATGC
AGAACGAGAATTGGTGTGAATGCCA
ACAACATACAATGATATGCACATTAAACGAAAAGACAGAATCCAAGCCGAAAAA
GAGTTTCCGTAACCCATTGGAAAGTCA
AAGAAACATAACTTGCAGATTGATCCGAACCTCCATAGATTAAAGCGAGTCAGT
GATTCCGGGACAAGGGTTATACCTGA
CTTAGTATCCACCTATCTTGCACAGTCCCTAATTATTATGTGACATCAACCC
ACCAAAGTCTCCCGCTGCGTTAAC
ACAAAGAAATCTTAAATGCAACTTCGAACCTTCGTATTGTTAAATGATAATGTC
AAGATAAAGTCAAAAGTATTCAAGAA
GTWSGTGTTCAACAGCGATACCGATAATTACCATCACACAAAGTATTCTACA
CCAAAACCTACCGTGGTCCAGGGTCGG
GGAATTACAAGGATGGTGCATTGATGAACAAATCAACAAGATACTTCC
AGTAATAAAAAGCCGCCACAAGAGA
AAGGTGTCGAACAATAACAGGTACAACAAGAGTTAAAGGGGTTAGTCCACG
AAAAGTTGACAAGAAACTTGTGAGTA
CTTGCTTCTGAGCAACGCAAGTATACCGAGGACTATTCCAATCTGAAATT
ACACAATAGCTTACAGTTAATGTT
TTTGAATACGTATCGTGGTGTGCCAAGAGACATGGAATAACTACTACAAG
TTTAAATTGATTGATTGCAACAATTG
AAGGCTTGCACATTGGAGGCAAATGAGCTTGAGGAGAGAAAATTGGATGCTG
CTAGACACCAACAGTGGCGGAAGAAGA
GAAGCTTNCCAAGAAAGATTGCGTTAGTATTGAAGATGAACGGACGAGTT
TGAGCAATTGCAAAAGCGAGTTGGTCA

F.9 5

GAGAAAGAAGGATTGGAAGAGAAAATTGCGTCGCCGTAGCTANANGCATCTT
TGANTGATAGTTTGAACTTGATAGCG
AAAATGACNATGAATCTTGAATTGNCCAAANTNAACAAGACTT

Fig 5 (cont'd)

>113g4 844bp in-house 1-844
ATAGAACTGTTGATATAACAACTCTCACTCCCAATTGTGACTTGAATAATAATAACCTATCACCTAGTAATCTT
ATCTTAACGTAAATCTCTGAAAGCACAACTCAATGTATAAAAGCATAAAGATAAAATCTTGGTAGGGTTAAAGTCATAAT
TATAATGAAACACAATTACTAAAAGGGATGGTATCAACAAATTATAGGCTAGGTAGAACCCATAGTGCCTGTCGGGAGTT
CGGGTAGTTGGAAAGTTGGGAAGGTTGGATAGTTGAGAAGGTTCCGTGGCTGATCTAAATTAAACAGAGAACGATAAT
AAATGTACAAAAAAACATTCAAGAATTAAACAAACCTTTATATATATATATTAAATGCTTGTCACTCAACTTGGCATTG
TGTGATGATGCTTCTGTAAATACCTTTAAGAACCGAGATTCACTATCTCAACTAAATTAAACCTTATACTTTT
GTTTGACATTCATAATGACACAAAAGATTGTGAATATTAGCTCAAGGGGATTCTACTCATTCATCTCAAAACA
CACATTCTTGTATCACCAATACCTTTGCTAACAGAGGAACAAAAAATTGACACGGCATGTCATTACCCATAGCACTA
TCACTACAAATCAAACGATTACAATAGTGGGAATGTCAAATCATGTATATTAAACACATTACACATAATTATTTCA
GGTACATAATACCAATATCTAAACCTCAAAATGGTACTGTACCTTAACTTCTCCTCATGTCTAGTTGAATATTAT
ACTTGCTAAATGTCAAATCATGTCTTCACACATTCCAGGTGT

Fig 6

>15c1 977bp in-house 1-977 b2

TTTTTTTACATATAGTGTCTTTTAAATTGAAACACAAAAAACAAAAAGTAAACTACTATCACCACCCA
 CCACCAACCAAAACATCATAGTGGAACTTAAATTGAAAGAATATTTAATACCTTAAATTAAACATACCTCAAAAGGAA
 TAGGAGTAAACCTTATGTAAATTAAATTAAAGCAAAAAAGGAAAGATTCAACAAATCTGTAA
 ATTAAATTAAATTACATTTCATCTTGAAGTGTATAGCTGTATAGCAGTAATATTAGCAATAATATTAAATTAAAC
 TTAAATTAAACAAATTAAATAGTAATAAACCGAATTAAACAAAAACAAAAAGGGGGGGGAAAGACAACCGAATAT
 AGAAGAAGAAAAAACACAAAGACCGGTAGTAGAATATCTGGCTTAAAGGCAATCTAAAGTACAGCAACACATAAT
 GCAGAAGACAAACCCATT&AACAAAGAATCATACCTCCAGAACGTGGTTGTGTGATACATACATAAGGTGTTGTG
 TTGTGTTGATAATATCCACCAACCTGCGTGTGTTGATAATATCCACCTGGAGGTGTTGCCATAAGCAT
 TATATCCTGTTGTTGATAGTGGCATGACCACCAACCCACTAAACATCCCTCGATCTGTGTTGTTGAGAA
 TATTGGGTTGTTGTTGTTGATACATAACTCTTGTGTTGATAGGTTGATTATTATAATTGGTGTGTTGACC
 ACTAACTTACCGAATATTCTCTTTGACATCTTATATTAAAGTGTAAATTGGTGTGTTGTTGCTAA
 GATTGAGTATAGAAGTGGAAATTAAACAAATTAACTGTGATAAGATGGATTAGCAATGATAATGAAC
 AAGTAAAGTGTGAAATGT

Fig 7

1 QQSVPQSCP NY5QTCORG MFSGGGGNG RY0000GVNA YGPPPPQQGY

51 YQQQPCGGGG YQQQPCQP NYVQQQPRSG GNDSCLNQCL AAIKVCTLD

131 MLE

>207g4 769bp in-house 1-769

GCAAGATCTAACTCCAGTTTGGTGTAAATGTTACACAAGCAAAATATAATCGAAAAGGCCAAATAATTCT
 CTTCTACAAATTACGAAAATGTTACATGTATGAAAAGCTTCTACTATCTCCAACTCTAGCAGTGAG
 ATGATACTGATATCTCCTATTAGGATACAGTTCTATTTAGTATAATAATCATGGAGATAAAATATATTAA
 TCGATGGAGTTAACGAGAAAACAAATACACCCATTGCGAGAAAATGAGACATTACAGAAAAAAACAGAAAAG
 ACAATTACTCATTCAATAATTCCACAAATAACAAAGAACAAACGTACTAACAAAAACATCACTAATTICA
 CTTGAAAATCTTACATACTCAACTTAAAGATTAATAAAAGCGATGCATATTCACTAGAATTAGTGTATACATA
 TCGAGGTGATTATGACCCAGGTAAACAAATTCTTACTAAATCTAGGAGTTGTTATATACAGTATTGTCTAAC
 CTGCTCTAACGTTACAGATAAGTTGTAATCGTTAGAAATAACAGAAGCTGTTGTTGACTTTGGTGGTGGTGG
 CAATTGAAATGATATATTGTTATCTCAAGTATAGCAAAATACAAGGGCAAAACGTCGAAACAAAACAAGAACTTGGATT
 GTCGCAATTCTCTCACCCCTTCAAGATGCTCTCGTGTATGTGATCAAT

Fig 8

>226c_af1 766bp in-house 1-766bp

AACGTAAATTGTTATATTACCAAGTAAACAGGGACCTCATTATCATTAGTTGCAATTCAATTACTCCAGAACAAAGA
 AACACAAGACTTGTGTTGCTTAAAGATAATATATAATCAGGATAAAAGAATTGGTAAAGAAAATTCA
 CAGGGACGGTAAATCATCTCTCTCCCTATAAACCAAAATCTTATATGTCCTAACGTTACTTATTAGAATTCCAGATT
 ATTACTTACAGTGAATCATTAAATTTAATTGAAAGCGAGTTAGCTCAATGTCTCAGACACAACCTGCCTTCAAG
 GCACCAACACAAAGCACCAAGCTCCATGGATCTGGTACAACTCCAAAGATCTCCAGCAAGATTGTTCAAG
 GTGGATATCATCATCATCAAAAGATAGCCAGTATATCCAGAAAAGCCCTCTCAAGAAGCAAAACATAGCACCAG
 AACCAAAATAACTAAACAAACAGTACCGAGCTAAACAAATAGGTACATCTGAACTCGTGCCTAAGTGTGGCT
 TCGAGTCATGATAATTCTAGTTGCAATTGCTATATTCTGATTCTAAATAACAAATAGTATGCAAAAT
 GTTACTCACAGTATAGAGGACATAATTAGAGGACATAGACGATGCTGAGATATACGATGCTGAGAAGGTACCCATAAA
 CATATATAAGTTCTAAATCATGCTAAACACATTAAATTATTG

Fig 9

Fig 10

>22g3 (5') 535bp in-house: 1-535

AGGTTCCAGTTACCAATTAGGAAGTGTGTTGCAAGCAGGGCTACCAAATATG
GGTGGCAACACATATGGTAGTAAGTGC
TACCAATGTGGGTGCAAAAAATTGCCAAGTAATTGTATGGCAATAACAGA
AGTGTGGCGGATTNACTGAGGAAT
CTTGGTGTGAAAAAAAGCAATAGCGACTACGCTACAANAGGCAATCNAT
TATTATTATAAAGTGGAGTTATATAT
ATNTTCTCGGGGGGGGGGGGNTTNGNNNTCCCCCCCCCCCCCCCCANNTT
TNTCGGCCCNCCACCNTCGGCCCTTC
TGGCTCCCCCNCNGGCCNCNGTAAATNCCTCCACCCNGGANAANGGNA
AANGGGAACNANNAAGGGGGGACNNN
NCACCCNATGGGAGGGAAAATCCNAANNTTNCCCCCCNCCNGCCNAAN
CCNCNTGGGGGGGGCAAANNCNGGGG
GCTNCNCNCCTNCCCCCCGCCNTNNCCNNNTNCNNCGANCTCTNNGNG
GC

Fig 11

>22g3 (3') 426bp in-house 1-426

CCCCCATATAACGTTGTCATAGCAACTCTGTCGCACCCATAGTGTGACTT
CTCGTGGTATAAAAAAAATTTC
TCCCAAAAAAAATCTTCTCCTTCCACCACTTTTCTTCTTCTTCTTCCCCATT
CCCTCCCAAATCCCTCATTTCCC
CATTTCCCTACCCCTCTGGCCCTGTATTCCAAAATTTCTCGGGGNTACGCC
CGAAGANAACCTCCCTCCCACCCACC
CATCTTGTCTGGNTTCGACCTTCGGCCTCANGGCTCCACCGTCGGGGNTCTTG
TATATTGTAGACTCCNGGAAAAAGG
GAAAAGGGGAGGAAGAAGGGGGAAAAAAAANGGAGGGNGAATCCTT
TTTNTTNTNCCCCCCTCTAAACCNAAA
CCCCNTNTGGGGGGTCNATTAGGGG

Fig 12

>35gK 1334bp in-house: 146-669 public: 1-145 PathoS q: 670-1334

ACAAACGTATAATCGACAGTTACTATATCTGCTGACTCAAAACCAATGCATTC
TTCAAGCGTGCTCTGTCGATTTCTAT
CATAACATCCACTTCCGGNGTAATCGGATTACTAAAGCCACAGAATCAAGGT
GAACATCAAGCTCAACTCTTCTTG
GTCCACGAATAATTAAATTGGTTTTSKKGSMAMKGCTTCTACRGTAGGTT
TGAATCTTCCAACATTGTCTTGCA
TAGAAAACMGCACCAAGACAAGAAACATGTCCACTCGACCATCAACYTSKGGGT
AWWGACAAAGTWAATCTGTCTGGATCCT
TTTCATCCAGTTCCCTGCATKGGAWACAAGTNTGTCCCGCACAGTTAAGACT
GTTTTATTTSKTGGTATTAGACTCA
TCAAGTTCCGAAGGAGAGGCATCATTARGGGWATAGACTCCGCTGAGTTAAT
ACTGGATAAAATCACTTATTCAGATT
ACTGACTTGTWCTTCAGTGACCTTATCAAAATCCTCAATGTACTCSGARGCGTW
TTCMCTCMATGTGAAGGCTTTAAAAA
GGGCAACRCTGGTTYCAAATGCTTCTGCRA GTTGTACKTGACAGAAAAAA
TCAAAAACYTTGAAAGATACCTCTT

Fig 13

CTAAGCTTTAAATCAATTCTTCTAATTTCATCATATAGCTTATGAC
 TTGCAAACCTCCTACATACCAT
 ATCCATTACAATGCTAGAAATGTCAATCTTCACTGACGATATAAAGGATGGAA
 GAACTTCAAATAATTATAAAACTCAG
 GATTGGCTGGTGTATCTGCTGCAGGAGCTCCAGATTATTGTCCATTGCTCAC
 TCCATGGACATACATTATAACGTCC
 ATCTTTTCCATTCTCAAATTCTCGGTGAAATAAAATTGTTGACGRWTTTA
 AACAGACGTACAATGTGAAAGATAA
 GATCATTAGCAGAGAGCAATTGAGACTCTGCTTGAAAGTTGATTGACACG
 TTTGTTGTAACATATTGTAGGTGGCT
 AAAAGATTGACTTWRGTAAAATGRAACTTATTAAACCTGGGCCCTCACATT
 ACATTTTCATCTTAAACAAAGKGGTT
 CAAAGKGGAACTTGGTTGGATCCYTTAWTGGAAWATTTCYCAGKRAATACTT
 TCAAAATCAACTCCAGGAGAGCCACAG
 TGATAATTGAATTGGATTAGATAAGCGGTTAAACTCCCAATTCACTTAC
 CAAACTCTGGTAAATGAAGGTTAAGT
 TTGTTGTCACCAACAAGTTACTAAAAACAGCCTGAGCATTGGAGGCA

Fig 13 (cont)

>36g2 (5') 520bp in-house: 1-520

CGTATAGAGAATAATCCGTTGAAATTGATTGTTCAATCATTATTGTATCTTTCC
 CTTTTTTTGTCTAACCTAAATGT
 TAGAATAATTAGAAATTGTTCAAAATATATTCACTGTTAACAAAAACAGAAAT
 GCTTCAATAAGATTGATTCTAATT
 ACTAATCGTTAATATTAGTTGGTGGGTTTATTATCGAAGATGTAGCATT
 ATTGTTATCNAATAGATAAAAGAAACT
 TGAATTAAATGGCNTAATTGTTGCAATAGTAAAAAGAAGAAAAGTGGTAAG
 GAGTGAGTGAATAATTGCCCCA
 ATTGAGTNGAAATCTACACCNAAGTTGGACNAAAAGTTTACTAAA
 ATCTGANAATCTNCCNTGAATAGAACCG
 ATCATCCNCATNTCCGATTCTGAGGANAGATAGTGGCCCCACCTCNTGGTG
 ATTAGAAGGAGCNCATGTTTACAA
 TATCTATATCCAGAATAACNTGTTGACCTNCNCNG

Fig 14

>36g2 (3') 472bp in-house: 1-472

CTCTATATATAGTGAATATAACATCAAATAATGTACAAAAAGTATAATAAA
 TTGATTTAGAAATGAGAAAAAGAAAA
 AACTTGAAGTAGTGAAGATATATTGTTGGCTATCTTCTGGTATGGCTCAAT
 TCAGCCAATCTGGATGAAAGGGTGG
 AGTTTGTGTTGTTATTGATTTGTAAGTACTTCCGGCTAGAAAGTTNA
 CAAACATGATTAATCTGATATANAT
 ATTGTTAACATTTGGTGTCCNTCTTAATCNCNAAGTTGGGNACTA
 TCTTCCNCNGAAATCTGTATATGT
 TGANTGANCCGNTCCATTCTGTTNANTTCNGANTTAGTTAAAACCTTTG
 TCCCAACCTTTGGGTTAGANTCN
 NCCCCANTGTTGCCNNAAATATTNCNCNCNCCTNCCCCCTTCCCCNTTTAC
 NAATGCACCAAGTAAGCG

Fig 15

>38g1 1348bp in-house: 183-940 PathoSeq: 1-182 / 941-1348

TCTCTGGTATAACTGCCTACCTCATCGCTACCCCGGATTTTTGGTATGA
 TCTACACGTCTCATCGCTACCCCCA
 GATTTTTCTGGTGCACGGACACGCCCTCCGGTCCGCACCGAAAACCGGGG
 TAATCTCCGTGGAGATACACATCCG
 CGGACACAAAATCAGATGAGCTACCACCGAAAATTCCGAAATTCAAAAAC
 AAAATCCCTAAAACAAACTATCCAGA
 NATTATTGCCATGCCCTGAGGATGAGTTAGTTTAATTITGAAAAATGTC
 CAAAACGGTTGTGCTGTAGGANG
 GGTAAAGAATTGCCATTCTGCCCTGGGTGGTCAGTCNAAAAAGANGTA
 TCACTCTGGTCNAACGGGAAACAACN
 NAAAATGGGATTAAAMTWATCTCCAGAMCAAACCTAGCTTMWWACACCCAY
 TTAGTTGACTSGYWRCCMAAMMCMAA
 TTTCATTGTTGGGANGGAATTARACCAAATTTTTGAAATT
 CGCTMAGTGYMAGAMCCSCAAAAG
 TCACCTTTCTGTTTCMTCYACGGCARARGCYCACCGGTTKYKTGGKGS
 MCRGCCMAATTGAWTTTGTGGTGSGC
 ACGKGAAAAACAGTTKGTAGTGGACACGTTTGCACTGTGAAACTGCGCT
 CGGAGGTACTATATGCGAAAGCAGAAA
 AGACAATTGCAAGAATACAGAGAGTTCTCTGGCTANGCAATGTGTTA
 AGGCCAAGTCGACGAGTGGGAGAGTC
 TGGAAAGTGATATACACATCACGACCTACTTATACGCTACGTTGGCATGGC
 GAGCCACTGTACGGTGGCAAGCCTGAA
 CAGTCCCACACCAGATATCTAACGATTCTGTATGGCACTGATGGGATTAG
 TGGATTACTAGCTGATAGCAAGTATT
 GAAAACAAAACCGACTCGGGGTATGCCTGGCAAGTAGCCGGAGAAAAT
 CTGTGACTTGCTGAGTGTAACTCCCT
 CCATGGTTGGCGATGTTGACGTGCGCGCAGTTCTGCGTATCACAGTCGA
 CGGACACCACACCGGGAGAATCTAA
 GAGGGCTATATGGATGTGGAACGGTTTGCTGCTGGTAAAACACTGGCGGG
 CGAGCCGACGTTCCACGGACACAGCAA
 TGTGTTGCAACAAATAAAACTGTACGGTTGAACGTGTTGGCTGCT
 CCTTCCAGTTCTGGCGGGAGAAGCT
 TGGCGCGGGAGACCACTACTACGTAGTTATCTGGTTGATCCTGCCAGTAGT
 CATATGCTGTCTCA

Fig 16

>60gK 990bp in-house: 445-752 public: 1-140/753-990 PathoSeq: 141-444

ATTACCGATCCGTCGGATTTAAAACCACAAAATTGCCTGCATTAGCAGAGCT
AGATATTTCATAGGGTCTATATATG
CAAAGATCTATTGAATGCACCCGTGAGGACACAATGTGATCACACGTACTGTT
CACAATGTATACGAGAATTTTTACTTC
GAGATAATAGATGTCCGCTTCTAAAACAGAGGTTTGAAAGTGGCTAAAAA
CGTGATCCATTGTTAGAAGAGATCGTC
ATTAGTTATGCCTCCCTTAGGCCTCATTGATTACGATTATTGGAGATTGAAAAG
GTGGAATCGAAGCAAGAGGTAGATCG

Fig 17

TGAGAAATCAGCCAATGAGTCAGCGCTGAATGGTAATAGAAATGTAAACAAAC
GATGTTGACGAAACTGTGCGCGTTAAAG
ATCAACTGAATGCAGATAAAACTAGGTGAAGAAAAAGGGCAAGCTAACATGG
GGAACAAAGTNAAACGAGCAGACTACTGA
AGTATTCTGTTGCTATCTGATGATGAAGAGAATGGTCTGATAGCCTAGTAAA
ATGTCCTATTGTTGAGAGAATGG
AATTAGATGTACTACAGGGAAAGCNTATTGACGACTGTCTAAGTGGAAAGAGC
ACGAAGAGGGACGCCCTACAGACATTAA
TCCCCAAAAGCCCAACGACCGAAGCAAATCACCTCCTTTTCCAACCAACAAT
AGATACCANAACNCCTCCCCCACCTA
CCAGTTNGCGTCNACAACCTCCACAGCAACTCCGACAACACTACATTGTTGAA
AGCAAACGTCTCATCTCCATCCCAAGT
GGCGCAAAGTACAGTAAACAAGGGCAAGCCATTACCTAAACTCGATATCAGCA
GCTTGAGTACTAAAAAAATAAAAGCCA
AGTTGAGTGTATGAAACTACCAACACAGGTAGTAGGAATGAAATGGAAGC
CAGATACTAGCATTACTATGTGATTAT
AATGCCAACCTTGACACCAATCATCCTGTA

Fig 17 (cont)

>64gB 627bp in-house: 1-627

TNCANCCATNCNCCAGGCNNNGCACCCNGCCNNCCCCNTNTTC
CCCCCTCTTNGCCTCNNGTG
GTGTTTGTGGTGTGACNAATAANATGGNTATCATTAGAANAGGACATTGCN
NCGGAAATGACTGTCGACAATAAAGAA
GCAAATATATAACAATGGATTATGAANGCTAGGATGGATTGAAAGTTATC
TGGGTTTATTCCAATGTAAGAAATTATT
TGTAATTGATATGGCTAATTATTTGCTNATNTATCACAAAAAAATGATTA
AGTCGAAATGAAATTGGCCTCCATA
TATAAAATTCTGACAGGAAGAGAAAATTGACNTGTTGCCNAAAAAAAA
AACTTACCCNCNTCNANTCNTGTNN
GACTTAACCCCCAAAANANANGCTGGCGCGNAAAAAAATAGGAGGGG
GCCGGNNGTTTAAATTNNNNCTT
GAATATGAACCCAANNTTGNNTTCNTTTNCCACNCCCCCTCAAATTNTAT
TCCATGTTCCCAAGANNAGGGNGNG
GGGGNGGTCCNNCTTTAAACCNCACCCCCGGTGGNGGGNCCGTNTNT
TTCCGGNGGGCNT

Fig 18

>8c_cp 890bp in-house: 287-890 public: 1-124/154-286 PathoSeq: 125-153

ATGCAATTCTCATCCGGTGTGCTTATCCGCTGTTGCTGGGTCCGCTTGGCTG
CTTACTCCAACCTCCACTGTTACTGG
CATTCAAACCACTGTGTACCATCACTCATGTGAAGAAAACAAATGTCACGG
AAACTGGAAGGTTACCAACTGGTGTAC
CACCGTCACTGAAGTTGACACTACGTACACCACCTACTGCCATTGTCAACCAC
TGAAGCTCCAGCTCCATCTACTGCTA
CTGATGTTCTACCACCGTTGTACCCATCACCTCATGTGAAGAAGACAAATGTC
ATGAAACCGCTGTACCAACCCGGTGT

Fig 19

ACCACTGTCACTGAAGGTACTACCATCTACACTACCTACTGCCATTGCCATCT
 ACTGAAGCTCCAGGTCCAGCTCCATC
 TACTGCTGAAGAATCTAACCCAGCTGAATCTTCCCCAGTTCCAACCACCGCTGC
 TGAATCTTCCCCAGCTAAAACACTG
 CTGCTGAATCTTCCCCAGCTCAAGAAACCACTCCAAAGACCGTTGCTGCTGAAT
 CTTCTTCAGCTGAAACTACTGCTCCA
 GCTGTCTCTACCGCTGAAGCCGGTGTGCTGCTAACGCTGTCCCAGTTGCTGCT
 GGTTTGTGGCTTGGCTGTTGTT
 TTAAGTTTATTAGAGCTTAAATCAAATTTACAAACAAAATTTCACTTCCC
 CCCTTCCCTTCTCATTCTCAAA
 AAAGGGTTATTTACTATTAAATTGATAAAATTATGGTTTCAATGTTAATTACCCCT
 TTCTTATAAACATTGGTATTATTA
 TTATCATCATTAGNTTATTTATTTGTGAGTTTTCGGNTTAAATTAATT
 TTGGATACATATTAAAAATTAT
 TTGGTACTAG

Fig 19 (cont)

>853 481bp in-house: 1-431
 CTAATATACTCGAGTTCTGGGGGTTGAAAAAAACGGTATTTTTGGACCCACGAGAAAAAAAGTGGATTTGGCGTGCA
 CGACCCGAAAAACGGAGAATTGGAAGAAATTGGGTAAAGTTTGAGAGAGTGTGGAGCAACAACATAAGA
 GAGGGTACGCCAAATTGTAATGGCAGGCGGGCAATGAAAGATGTGTTGCAAAAGATGGAGTTTGAGGGTTG
 GTGGCTGAGATATTGGCAGTTTAAAGACCCCATGTTTGGGTAGCCGTGGGTTAAGACCACATTTTTTGTAA
 GAGACCCGAGAAAAGAGAGCACACATACAAAATCAAGACCGAGAAAAGAGAGCACACATTAGAGCACATTTGGT
 ACCACACACTTTAAGAGCACAGAAAAAGAGCACTTATTCAGACCCATGTTGGTAGCACACACTTTAGAGCA
 C

Fig 20

>66g4 579bp in-house: 1-579

CCCCGTTAACCACTCTAGGGTATACCATTCTACTGACTGAATAACTGGTTAG
 TCGATTGTTGAGAAAAGTGAC
 CACCTAGTTTCTGCCAACATTTGCGATGAGCCGTCACGCCTGCTTT
 TTCTACCCCACGTTAACAACTCTG
 CCAGTCATTCCCTAGCCAAATAACTTAACTGACTCACAACTCTAACACTGACTC
 GTGCCCCCTGTTAAACTCTAAATT
 ACTTCACAGAGCCTTACTACCTAAATTARGRTTWSKAKKGTTCTGTTTT
 TTGCAAATCACCTGACTYGT
 TTTCAGCCAGGTTCTGTTAAATCTGACCAAAAAATTACRACTCCTATWT
 TTAAACTCYAAAWACAATTAAAC
 TCAATTCAAGACAAGTCCTCTGCTCATTCTGAGTCTCTATTGTCTTGTACT
 TTTGTGTGTGACTATTTCATGAT
 CACCCCGTTCTGCTTCAACTTTCTCAAAATCAAGCCAAAAA
 AACACACCTTAACACCTATAACAA
 CGCAAACCTATTCAAAACA

Fig 21

>NDI (17c_cp) 807bp in-house: 1-614 PathoSeq: 615-807

AACCTATTCCATAATGTTACTAGATCATTGATTAAAGGTGGTGGCAGACTTGC
 TACTACCAGATCATTGGTCAACAACT
 CTACTAGTTGGTTAAAAATCAATTAAAGAAATATTCAACATCAACTCCTC
 CTAAGGGTGCCTAACTCAAAATCTCG
 ACAATTGGTAAATATTCAAGATAACACTTTTACACTGCTGTGATATCGGTATT
 GGTCTGCCGGTTGATCGGTACAA
 AATTACGAAGAGTCTCAACCTGTTGATCAAGTGAAACAAACACCATTGTTCC
 TAATGGTAAAAAAAGAAAATTTAG
 TTATTTGGTTCTGGTGGGTGCTATTCAATTGAAAAACTGGATACCA
 CCTTGTATAATGTTGNTATTGTCCTC
 CCAAGAAACTATTCTTCAACCCATTGTTACCATCTGTTCTACCGGTACTG
 TTGAATTGAGATCTATTATGAACC
 TGTCAAGATCAGTCACCAGAAGATGCCCTGGCAAGTTATTACCTGAAGCAGA
 AGCTACAAATATNAACCCCTAAACTA
 ATGAGTTGACACTAACAAAGTACTACTGTCCGTTGGTCATTCTGGTAAAAAA
 TACTTCCTCTAAATCAACTGTT
 CCGAATACACTGGGGTTGAAGAAATCACTACCACCTGAATTATGACTATT
 GTTGTGGTGGTGGCTCAAACAAATN
 CTANTTTCGGNAATCCTGGGAGNCGNTGAGGAANTCAACCCCTTTTGAA
 AGAANGNCCAGTGGANGCCNTCTGCN
 AATTAGA

Fig 22

>HOL1 (409c5) part2 762bp PathoSeq: 1-762

GATCAGAATAATGAGGACTTTACCTGGAACACTCAATATCTATTCTTGGAA
 GTTGAECTCTGAAGATGAAAACGTGAG
 TCATTACGATGCTTCCAGTCGACCAAAAGTGAAAACAAAAGGCAATATAATCC
 TCTTCCCACAACCATCGAATTGATGCA
 ATGATCCATTAAATTGGAGTAAATGGAGAAAGCTAAGTAACTTTTTATTGTCA
 TTTTATTACTGCTTTACAGCAGCT
 ACTTCAAATGACGCTGGATCAATTCAAGATTCAACTTAATGAAAAATATGGAAT
 TAGTTACGACGCAATGAATACAGGGC
 AGGCCTTTATTTGGGTATTGGATGGGTACTTTCTTTAACACCTGCTTCG
 TCGTTATATGGTCGAAAATAACAT
 ACTTTATATGTATCTTCTGGTTATTAGGCCTGTTGGTTGCCTTGGTAA
 AAGCACTCCGACTCAATTGGTCG
 CAATTGTTGTTGGTATTAGTGAGAGTTGTGCTGAAGCTCAAGTACAATTAAAGT
 TTATCAGAACTTTATTTGCCATAA
 CCTTGGTTCTGTGCTTACGCTCTATATTGTTGCAACTCCGTAGGTACTTACTTA
 GGACCTTTAATTGCAACGCTTATTG
 TTCAAAACATTGGTTAGATGGGTGGATTGCAGCAATTATTAGTGGTGC
 CATTATTGTTGTAATTGTTTTGT
 TTAGATGAAACCTATTGATCGAGCAAAGTTACCAAGCCA

Fig 23

>GAL2 (360c6) 1004bp in-house: 625-1004 PathoSeq: 1-624

TCCATTTCCTTTCTCTTTCTACATCATCCTCACANCAATTCAAATATG
 TCTCAAGACAACGCTCTCATCAACAT
 CTACAGCTGAGGCTGAAATAATGAAATCAAAGTCAAAGATGAATTCCACAA
 GAAGAACAGCTCATACTAGTTAGAA
 GATAAACCAAGTGAAGTCATACATTGGTATCATCATTATGTGTTCCATTGCC
 TTTGGTGGTTTGTTCGGTTTCGA
 TACTGGTACCATTCGGTTTATTAAATATGTCTGACTTTAGAAAGATTGGT
 GGTACTAAAGCTGACGGTACTCTT
 ACTTTCCAATGTCAGAACTGGTTAATGATTGGTTGTTAACGCTGGTGTG
 CCATTGGTGMWTTATYCTTGTCTYAAA
 GTCGGTGATATGTATGGTAGAAGAGTTGGTATCATGACTGCTATGATTGYCTAT
 ATTGTTGGTATTATTGTTCAAATTG
 TTCTCAACATGCTTGGTATCAAGTCATGATTGGTAGAATTATYACTGGTCTTGC
 CGTGGTATGTTATCAGTTTATGTC

Fig 24

CTTGTTCAATTCCGAGGTTCTCCAAAACATTGAGAGGTACTTGGTGTGCTG
 TTTCCAATTGATGATTACCTGGGT
 ATCTTCNTGGNTATTGGCTACCTATGGTACTAAGAGTTACTCAGACTCTAGAC
 AATGGAGAAATTCCATTAGGTTATGT
 TTCGCCTGGGCTTATGTTGGTGTGGTATGGTAAATGCCAGAATCTCCA
 CGTTACCTTGTGGTAAAGACAGAAT
 TGAAGATGCTAAAATGTCATTGCCAAAACAAACAAGGTTCTCCAGAGGACC
 CAGCATTATACCGTGAACCTCAATTAA
 TCCAAGCTGGTGTGAAAGAGAAAGATTGGCCGGTAAAGCATCTGGGTACT
 TTATTCAATGGTAAACCAAGAATCTT
 GAAAGAGTTATTGGTGTCAAGCCTACAACAATT

Fig 24 (cont)

>KGD2 (98c_cp) 334bp in-house: 139-334 public: 1-138

TTCTAACAAACACATTTCTGGATCTCAATCAATTCTGATGGTCTTAAG
 AAAATAACAGCTTACGACCGTCAA
 CTACTCTGGTCGTAAAGTCAATGCTAAGTACATCATTGGTCTAGAAACGATT
 GTCCGTTAACAGNAATTGGTCTTNT
 TTAAAANTGTGAAACCAAATACGGNAGTTAANGCATTATAATTGGGT
 ACAGTATAATGATCCAATAACACNGNC
 ATTAAAATAGTGAAGAACCNCCGGTCATATCTTACAAAGTCAATTACNAT
 TTCTGGCTTNTTACNCAAATTANANA
 TTTCCTTTNAATA

Fig 25

>RNR1 (38) 2562bp in-house: 1-2562

ATGTATGTTATAAGAGAGATGGCCGTAAAGAGCCAGTACGTTCGACAAAAT
 CACTGCCAGAGTTCAAAGATTATGTTA
 CGGTTTGAATCCAAACCACGGTGAACCAGTTGCTATTACCCAAAAGTTATATC
 AGGTGTTTACCAAGGGGGTTACTACTA
 TTGAGTTGGACAACCTGGCTGCAGAAATTGCTGCTACAATGACAACAATTCAC
 CCAGATTACGCTGCTTAGCCGCTAGA
 ATTGCCGTATCAAATTACATAAGCAAACCAACAGTATTCAAAGTGTCA
 TAAGGATTATATGAATACTTAAATCC
 TAAGACTGGGTACACTCTCCTATGATTCCAAGGAAACCTACGACATCATTAT
 GGAACACGAAGATGAATTAAACTCAG

Fig 26

CCATTGTTACGACAGAGATTTAACTACAATTATTTGGGTTCAAGACTTGG
 AAAGATCATATTGTTACGTATCAAC
 GGTAAGGTTGCTGAAAGACCACAACATTGATCATGAGGGTTGCTGTCGGTAT
 TCACGGTAATGATATACCAAGGGTCAT
 TGAAACCTATAACTTGATGTCATAAGATTCTCACCATGGTTCTCCTTGTAA
 TTTAACGCTGGTACACCAAGACCAC
 AAATGTCCTCATGTTCTGCTTGATGAAGGATGATTCTATTGAAGGTATT
 ACGACACTTGAATCGTGTGCTTG
 ATCTCAAAAAGTGCTGGAGGAATCGGTTACACATCCACAAACATTGTTCTACC
 GGTGCTTACATTGCTGGTACCAATGG
 TACTCTAATGGTATTATTCAATGGTAAGAGTATTCAATAACACTGCACGTTA
 TGTCGACCAAGGTGGTACCAAGAGAC
 CTGGTGCTTGCCTTGACTTAGAACCATGGCACAGTGACATTGATTCA
 TTGATATTAGAAAGAACATCGGTAAA
 GAAGAAATCAGAGCCAGAGATTGTTCCAGCTTGGAATTCCAGATTGTT
 ATGAAAAGAGTTGAACAAAATGGTGA
 CTGGACTTATTCTCACCAAATGAGGCCAGGCTGGCTGATGTTATGGTGA
 CGAATTGAAAGAATTATACACCAAAT
 ACGAAAAAAGAAAACCGTGGTAGACAGACCATCAAAGCTCAAAATTGTT
 TGCTATTGGGAGCCAAACTGAAACA
 GGTACCCCATTTATGTTATATAAGATTGTAACAACAAATCCAACCAAAA
 GAACCTGGGTATTATCAAATCTCCAA
 CTTGTGTTGTGAAAATTGTTGAATATTCTGCTCCAGATGAAGTTGCTGTTGAA
 CTTGGCTTCCATTGCTTGCATCAT
 TTGTTGAAAATGATGAAAAAAGTACTGGTACAACATTGACAAATTACATCAG
 GTCACTAAGGTTGTCACCGTAACCTG
 AACAGAGTTATTGACCGTAACCATACCCAGTCCCAGAAGCTGAAAGATCAA
 CATGAGACACAGACCAATTGTTGGG
 TGTTCAAGGTTGGCTGATGCTTATGGAATTGAGATTACCAATTGACTCTCA
 AGAAGCTAGAGAATTGAAACATTCAA
 TTTTGAGACTATCTACCATGCTGCTGTTGAAGCTCAATTGAATTGGCTAAAG
 AAGAAGGTGCCTACGAAACCTATCCA
 GGTTCTCCAGCCTCTCAAGGTTATTACAATTGATTGGAACAGAAAACCA
 ACTGAATTATGGGATTGGGATACATT
 AAAACAAAGATTGGCCAACATGGTATGAGAAACTCCTGTTGGCTGCACCAA
 TGCCTACTGCTTCCACATCAAATT
 TGGGTAAACAATGAATGTTGAACCATACACTTCTAACATTACTCTAGAAGAG
 TATTAGCTGGAGAATTCCAATTGTC
 AATCCATATTATTGAGGACTGGTTGATTGGGTGCTGGAACGACGCTATG
 AAAAGTAGTATTATTGCTAACATGG
 TTCTATCCAAGCCTTACCAACATCCCTGATGAAATCAAGGCATTGACAAAA
 CTGTCGGAAATCTCACAAAAACATA
 TTATCGACATGGCTGCTGATAGAGCAGCATTATTGATCAATCTCAATCATTAA
 ACATTACATCAAAGATCCAACAATG
 GGTAAATTAAACCAAGTATGCACTTCTACGGTTGGAAGAAAGGTTAAAGACTGG
 TATGTACTACTTAAGAACACAAGCTGC
 CAGTGTGCTATTCAATTACCATGATCAAAGATTGCTGAGACTGCCGGTCA
 TACGGTTGCAAACATTGGACAAATTAA

Fig 26 (cont)

ACATTAAGAAATATGTTAACAAAGGAAGAGTTGAGAGTGAGAATACCAAGTGAT
 GCTCCATACAAGTCACCATCAACCGAA
 CCAACCTCATTAGAAAGTTAGTTGCTGATTGAAAATAAAAGATGAAGGTGA
 AAAGCCAGCTGAAGACAAAACCATTGA
 AGAACTCGAAAATGACATTATAGTGCCTAAAGTTATCGCATGTGCTATTGATA
 ATCCAGAACATTGTACAATGTGTTCTG
 GT

Fig 26 (cont)

>SAM2 (36) 1155bp in-house: 1-1155

ATGACTACTTCCAAGGAAACTTCTTTCACTTCAGAATCCGTTGGTGAAGGT
 CACCCAGATAAGATTGTGACCAAGT
 CTCCGATGCCATTAGATGCTTGTAGCTGTTGATCCATTGTCAAAAGTTGCT
 TGTGAAACTGCTGCCAAACCGGTA
 TGATTATGGTTTTGGTCAAATTACCACTAAAGCTCAATTGGATTATCAAAAAAA
 TCATTAGAGACACCAATTAAACACATT
 GGTTACGACGATTCTGAAAAAGTTTGTGATTACAAGACTTGTAAACGTCTGGTT
 GCAATTGAACAACAATCTCAGATAT
 TGCTCAAGGTTACATTACGAAAAAGCTTGGAAAGAGTTGGGTGCTGGTGTAC
 AAGGTATTATGTTGGTTATGCCACCG
 ATGAAACCGATGAAAAATTGCCATTGACCATTATTGGCCACAAATTGAAT
 GCTGCCTTGGCTTCTGCCAGAAGATCA
 GGTTCCCTGCCATGGTGAAGACCAAGTACCAAGTCAACATGCCAGAAGAAATCACCA
 TGAAAAAGATGGTGGTGCAGTTATCCC
 AAAAAGAGTCGACACAATTGTTATTCCACTCAACATGCCAGAAGAAATCACCA
 CCGAAAATTGAGAAAAGAAATTATTG
 AACATATCATCAAGCAAGTCATCCCAGAACATTATTAGACGACAAAACATATC
 TACCACATTGCCATCAGGCAGATT
 GTCATTTGGTGGTCCCCAAGGTGATGCTGGTTGACTGGTAGAAAGATCATTGTT
 GACACCTATGGTGGTGGTGGTGCACA
 TGGTGGTGGTGCCTTCTCAGGCAAGGATTCTCAAAGTTGATAGGTCTGCTGC
 TTATGCCGCTCGGTGGTTGCTAAGT
 CGTTGGTGAACGCCGGATTGCCAAAAGGCCCTGGTGCAGTTCTCCTATGCTA
 TTGGGGTTGCTGAACCCACCAAGCATT
 TATATAGACACCTATGGGACATCTAAATTGAGCACCGAAGCCCTGTAGAAAT
 TATCAAGAATAATTGACTTACGCC
 TGGCGTAATTGAAAAGAATTAGATTGGCTCGTCTATTATTAAAACCGC
 TTCTTACGGACATTACTAACCAAG
 AAAATTCTTGGAACCAACCAAAAAATTAAAAATT

Fig 27

>135g 859bp in-hous 1-859
CGTGCATAATTATCTTAAACCGTAGATAAGCAAAAATTTATCTATGAAATGTCAGCGATAAGAAGAAAGAAATCAG
GTACCCACGAGGAGTGTGAGAAAACAACCTCGTAAATTAAATGAAATCTAGTTCTCTATACTTGAATAATTGGAGT
TTTCTGAAAAGACACCTGTTCCAGTTCAAAATTAAACAAGAATGTGAAAAGAAATAAATTGGATTTATCTAGGCTGTT
AATAACTCCAGGAAAACCTCAATTTCGTAAATTGGCAACTTGTCCGAGTGGCTAAACGAGAAAAGATTAGAAATCTTTGGGCT
TTGCCCCGGCAGGTTCGAGTCCTGCAGTTGTCGTTATTTTTGTTTACTCTCTATTTAAATTAAACTAAATCAA
CTGAAACTGGAGTACCTGCCATGATATGAGTAAATACTTTTTGATATTAAAAACTATATAAAACTCCCTATTIATTT
TTAATTAAACCCAGATAATTGTCCTAAATAGTTTTGTTGAACTTATTGCTTTGATGAACTTGTGATT
TTTCCAATTTCATACCTCTCTAGTTGGCACATCAGTGGCTATTGAATAATTCTGATCTGAAAGTGTACCCAGATGTATT
CTGACAAAACCTCCACACGGACCCACTCAATAGCATTAGATAATTGATTTAAAGTTCACOGAATATCGAATAATCTT
TATTGGCCATCTCATCTCTTGCATAATTCTTAAACCGTACTTTCTCAACCTTATTATCCCTCTAGATAAC
TCTTCCAAATCTTCAGGTTCAATTACCTTAAACCATCAATGAACAACTAGGGCAAAAC

Fig 28

328c2

X X fs
 = = =
 1 MSIITVTFPKS PCTYXRAFAF GIEBLEFSQGS SSUGAIKAA LAVFVFSVDN

X X X R
 = = = =
 51 QDFVLIADLA KYNTYYPSSYQ LIVKLVKCAN IEKSQILKTD KDLNKEFL
 101 DLISBEADTRI DLFYVISLPLV YSRRIENKCFY YVLRREPQPK VSKAPTQEXP
 151 ASVVAEEEDD DNLDDDEEDE YDQDMDEEND NSGELSKGYK HMMKDNPKYI
 201 NDDRVТИGQV FHQYQDLPST FLTHSLFNSI NEMGKLNYYK NFGVSGITPL
 251 PNSKLS7AER ELVLMANNYN DMHINEKTES XPKKSFRKPI GKSXKHNLQI

fs t
 - -
 301 DFNSICLSES VPGQGFIPD FSIMHLLCKV? NYVYTSNHQS LPLSTNTKNL

X
 =
 351 NATSNSSYLF HCVVCKSKS IOKLIVFNSDT DNYHHTK/FY TKTYRGPGSG
 401 NYKDGALMEK ENKCHLSSAK XPFHKRKVSN MNRYRNLKLG LVHEKFDKNP
 451 VEYLLSEQRK VTEQYENLEI LHNSLQFNLV LNTYRGVAQE TWNNYYKPL

X fs
 = =
 501 IDFEQLKALQ MEANEEERK CDAARHQWQ ESEKURQERL RUVFEDEPNE

X X x x xx
 = = = = = =
 551 FEQLQSEFGQ RXXDLEEEERK ARQLEASLSD SFADSEENDG ESELAQIQDD

missing sequence

===== 601 FESSANALKT KPEAYYVHLI NPAPBPCPTE TPQDUNNKF SLPTVYPEI

missing sequence

===== 651 PNPLPLERGV VPESKBBLPP IKKAIHVTT VPERPNKEYL TRNRCYPLAN

missing

===== 701 ANSGWKG

Fig 29

15c1

f8 s

1 CQSYYVZQSQ7 NYSQQTQDRG NFSGGGGGHG HYQQQQGTYA YGPPPPQGGY

ambiguities

x w w w

* * * * =

51 YQQQPGGGGG YYYQCCQGQ7 NYQQQQFSG GATGCLYGL AALCVCCTLG

amb

==

101 MLF

Fig 30

222g8

s

e e

=

= =

1 MRRREIERRX JQHXRQCRK EHEANRDIRI QQLSEQDSRS NQTKXEEEXVF

51 KARSTNSGA DETGLMSDKR FDDSAVSPDY LFZENLWNKP NHPTDNHETK

101 KYTENVVEND OSPNDTSAY NGSTHDETNI QNEIQIPEEND BYVPMKATS

K D VR f8 C

* * * * =

151 SVNNTTIPAG BKHESLSTSE NYARYFETAD VGVXGLESFY XAQTRNIWKG

P

=

201 QVSDNPWATV FTMOKRLES PEGKLLCRDQ

Fig 31

book

FIG

1 ITDTSOFKTT KLPALAMELDI LKRCYICKDL LNAZVPTGCD HTYCSQIRE

51 FDLRDNRCPD CKEVFEGL KRDPLLEHIV ISYASLRPHL LRLLEIEKVE

101 SKQEVCREYS ANESALMGR NYMNVDET/ RVKDLQNAOK LGEEENGQAH

G fs

= =

151 WEQVNEQTOB VILLESDOBS NGSDSLVKC? ICFERNELDV LGONHIDCL

X

=

G G X ambiguities

= = - - - - -

201 SGKSTKRTPT CILSPKAKRP KQITSFFKPT ICKTKTPSPPT SKASTTPATC

S Q N I K M

= = = = = =

251 PTTILLLKALP/ ASFSPVXQST VHKGKFLPKL DFEGLS3QKI KAVLSDLKLP

301 TTGSRNEMEY PVLYYYIVW AYLDSNHFW

Fig 32

sec 62

G G fs D A

fs

1 XQFSSAVLVS AVAGSALAY SNTVTDQI T/TITSCDE NMCHSTEVT

51 GVTMTEVIT TYTYCPLST TEPAPSTAT DVSTTWTIT SCEDBKCHET

101 AVTGSVTTVT EGTTIYTTC PLSPEARGP APSTABESEXP AESSPVPTA

151 AESSPAKTTA AESSPAQETT PKTVAEESSE ABTTAPAVST AEAGAAAANAV

201 PVAAGLIALA ALF

Fig 33

17c cd

1 PPKVAKSKES TIGKIFRYTF YTAVISVIGS AGLIGYFIVE ESUPVGDQVKQ

X

=

51 TQUPFNGEKK KTLVILGSGW GAISILKNAO TTLYNNVIVS PRNYFLFTPL

f6 X f8 f8
= = = =

101 LPSVPTGTVE LPSIIEPVRS VTRRCPGQUI YLEAFATNIN PKTNELTLMQ

R N X X
= = = =

151 STTVVSGHSG KITSESKSTV AEYTGVEZIT TTLNYDYLVV GVGAQTLIF

X X X XX XX X
= = = == ==

201 GNPGRMRKF NPFERIICSG SHLQIR

Fig 34

40965 DANI

1 DQNNEDFPG TNIYSLVED SEDENVSHYD ASSRPKVTK GNIIILFPCPS

51 NSCNDPLNWS KTRALSNPFI VIFITAFIAA TSNDAGSIQD SLNEKYGISY

101 DAMNTGACVL FLOICKGTFF LTPASSLYGR KITYPICPL OLLGAVWFAL

151 VKSTSOSINS QUPGISESC AEAQVQLSLS ELYPAHNLSG VLTSYIVATS

201 VGTYLGPLIA APIVGNIGFA WVGWIARIIS GALLFVIVPC LDETYFDRAK

251 FTKP

Fig 35

280c9

1 DNVSSSTSTAE KDNNEIKVCD ETPQESQAHY SLEOKPVSAY DQIIIMCFLI
 51 AFCCFVFCFD TGTISGFINK SDPLERFGGT KADGTLYFSN VRTQLMIGLF
 X X X
 101 NAGCAIGALF LSKNQDMYCR RUGINTAMIV YIVGIIUQIA SQHAWYQVMI
 ambiguities
 X
 151 GRIITGLAVG NL5VLCFLFI SEVSPKHLRG TLVCCFQLMI TLQISFLGYCT
 fs
 201 T/GIKSYSDS RQKRIPLGLC FAWALCLVAG MVRMPESTRY LVGDRIEDA
 PR
 251 KMSLAKTNKV SFEDPALLYRE LQLIQAGVER ERLACKASWG TLFINGKTKLF
 IV missing sequence
 301 ERVNLGVMLQ ALQGPWQKN LFPSYLTTSXP N

Fig 36

280c9

missing sequence
 1 NAFVSGCITE PLWVUDATIVE VQ2EITKME GDAPAGGASA SEADAKXEEA
 missing sequence
 51 PEKAKESAP ALQPKHESTK KEEPKCESKP APNKEESKK6 TQSTTSAPTF
 missing sequence
 101 TWFSPNEERV KMKHPLRCA ERLKESQNTA ASLTTNEVD MSNLMDFRKK
 missing sequence
 151 YKDEFIEKTC IKLTFMCAFS KASALALKEI PAVNLLAENI DCUVFKDYAD
 missing sequence X XX XX NK
 201 ISIAVAT7KG LUTP7V7RVAE SLSILGIEKE ISNLGKWARD GKLTLEDMTG
 S X XX Z C X X X F X F X IX
 251 GQFTISNGGV FGSLVYETPQI NYPTQAVLGL HGWKRPVTV NGQIVSRPMN
 301 YIALTYD4KV VDQSPRVIPL RTIKELIESP RKVLL

Fig 37



(19)

Europäisches Patentamt
Europ an Pat nt Office
Office eur péen des br vets

(11)

EP 0 982 401 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
01.03.2000 Bulletin 2000/09

(51) Int. Cl. ':
**C12N 15/31, C07K 14/40,
A61K 31/70, A61K 38/16,
C07K 16/14, G01N 33/50,
C12Q 1/68**

(21) Application number: 98310694.9

(22) Date of filing: 23.12.1998

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE

Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 14.08.1998 GB 9817796

(71) Applicant: JANSSEN PHARMACEUTICA N.V.
2340 Beerse (BE)

(72) Inventors:
, Contreras, Roland Henri
9000 Gent (BE)
, Nelissen, Bart

2340 Beerse (BE)
, De Backer, Marianne Denise
2340 Beerse (BE)
, Luyten, Walter Herman Maria Louis
2340 Beerse (BE)
, Viaene, Jasmine Elsa
9000 Gent (BE)
, Logghe, Marc George
9000 Gent (BE)

(74) Representative:
Baldock, Sharon Claire et al
BOULT WADE TENNANT,
27 Furnival Street
London EC4A 1PQ (GB)

(54) Drug targets in *Candida albicans*

(57) Nucleic acid molecules encoding polypeptides that are critical for survival and growth of the yeast *Candida albicans* are disclosed. Also provided are methods of identifying compounds which selectively modulate expression or activity of such polypeptides comprising the steps of (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid

molecule according to the invention which mutation results in overexpression or underexpression of said polypeptides in addition to contacting one or more wild type *Candida albicans* cells with said compound, and (b) monitoring the growth and/or activity of said mutated cell compared to said wild type; wherein differential growth or activity of said one or more mutated *Candida* cells is indicative of selective action of said compound on a polypeptide or another polypeptide in the same or a parallel pathway.

E P 0 9 8 2 4 0 1 A 2

1 M^WVYKEDGRK E⁷YRF⁵KITA R^YQRLCYGLN P^WHVEPVAIT QKVISGV^YQG
 31 V^YTIELDN²A A²LAATMOTI H^YPO^YAVLAA² R^YAVSNLHKQT TKQYSKVSKD
 101 LY^YEYINPKTG L^YHSPMIS²KET YD^YILM²HED²E LNSAIVYDRD F^YNYNYFC^YFKT
 151 S^YERSYLLRIN G^YVAER²PQH² IMP^YVAVG²NG ND^YPRVIET²Y NLMSQRFFTH
 201 G^YSPCLF²YAGT F^YFPQ²MSSCFL LAMKDD²SIEG Y^YDTLKSCAL ISKSAGGI²GL
 251 H^YHNIR²STGA Y^YASTNGTSN G^YIPM²YRVFN NTARYVD²GG N^YK²PGAFALY
 301 L^YEPW²HSD²IFD F^YDI²RK²NHCK E²IR²ARDLPP ALW²IPDLMK R^YZQNGDW²TL
 351 F^YSPNEAPGLA D^YVG²DE²PEEL Y^YTKY²EKENRG P^YQTIKAQKLW Y^YAI²LG²AY²TET
 401 G^YTF²FMLY²ADS C^YX²JK²SNQK²ML G^YI²K²SSN²LCC E²IVEY²EP²E VAVCN²L²ASIA
 451 L^YPSF²VEN²DEK S^YT²Y²N²F²KLH Q^YTKV²Y²TRNL NR^YTD²R²N²Y²P V^YPEAER²SN²MR
 501 H^YR²PIALGVQG L^YACAFM²ELRL P^YFDSQ²E²AR²EL NIQ²IFET²YH AAV²EAS²IELA
 551 K^YEGAY²ET²Y²P G^YSP²A²S²G²GLL²Q F^YDL²W²NR²K²FT²E L^YWD²WD²TLXQ²D L^YA²X²HGN²RN²SL
 601 L^YV²APM²PT²AST S^YQ²I²GN²NECF S^YP²Y²TS²NIY²SR R^YV²L²AGE²F²IV N^YP²Y²LL²EDL²VD
 651 L^YG²W²ND²AN²Y²S S^YI²IA²NG²SI²Q AL^YW²I²D²E²IK AL^YI²K²T²V²WE²IS Q^YK²H²I²D²MA²AD
 701 R^YAA²F²D²Q²S²Q²S L^YW²H²I²K²D²PT²Y² G^YKL²T²M²H²Y² G^YW²K²GL²K²T²HY² Y^YIR²T²Q²A²S²A²A
 751 I^YQ²FT²I²D²K²IA² E²T²A²G²T²V²AI² E²W²LN²I²K²Y²ON² K^YGR²V²E²NT²S D^YA²P²Y²K²F²STE
 801 P^YTS²LESS²V²A²C L^YX²I²O²E²G²E²K² F^YA²D²K²T²E²E² L^YND²I²Y²S²A²V²I²A² C^YA²I²S²N²P²E²C²T²
 851 M^YCSG

Fig 38

1 MTTSK2TFLF TS2SV3ZCHF DKICDQVSDA ILDACLAVDF LSKVACETAA
51 KTGM2N/FG2 ITTYAQLDYQ KIIADTIKHI GYDDSEZGFD YATCNV2VAI
101 EQQSPDIAQJ LHYEKALBL GACD2GIMFG YATDETDEKL B2TILLAKL
151 XAALASRRS GSLPWLR9CT KTQVTTSEYK DGGAVIPKRV CTIVISTQKA
201 EEITTENLRK EEEHIIKQV IPEHLLJDKT IYKIQPSGRF VIGGPQGCAG
251 LTGRKIIVTT YGGWIAHGGG AFSCKDPSKV DRSAAAYAARN VAKSLVUTAGL
301 AKRALVQFSY A2GVR3PTSI YICTYGT5KL STEALVEIIK NNFDLRPGVI
351 VKEELDLARPI YFKTASVYGHF TNQENSWEQP KKLKF

Fig 39